Development of a chick embryo model to study important regulatory domains of human genes implicated in Motor Neurone Disease

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Kejhal Khursheed

September 2016

Declaration

The data in this thesis is a result of my own work. The material collected for this thesis has not been presented, nor is currently being presented, either wholly or in part for any other degree or other qualification. All of the research, unless otherwise stated, was performed in the Department of Pharmacology, Institute of Translational Medicine, University of Liverpool. All other parties involved in the research presented here, and the nature of their contribution, are listed in the Acknowledgements section of this thesis.

Acknowledgements

I am very grateful to John Quinn for giving me the opportunity, and his tremendous support throughout the PhD. I would like to thank Diana Moss for great support, guide, and advice for academic and practical life. I would also like to thank Jill Bubb for her support, and always being like a harbours for experimental problems. I have been fortunate to have such a lovely man thank you Ali for your sport and love while I have been slightly mad writing this thesis. I am very great full for a lot of hug and love from my little princess Cenaria and young Yousif. Support and assistance also come from department staff, other members of the lab, both past and present start from, Rachel Carter, Sokratis Theocharatos, Palmer Megan, Christine Cashman, Christine McNamee**,** and Michael Lyons. Thanks also to Neurobiology group: Kate Hadley**,** Thomas Wilm**,** Veridiana Pessoa, Alix Warburton, Olympia Gianfrancesco, Abigail Savage, Paul Myers, Kimberley Billingsley and Maurizio Manca.

I would like to thank the Khursheed family, to my dad and mum souls, all my sisters and brothers for their support and encouragement.

Publications

- Khursheed, K., Wilm, T.P., Cashman, C., Quinn, J.P., Bubb, V.J. and Moss, D.J., 2015. Characterisation of multiple regulatory domains spanning the major transcriptional start site of the FUS gene, a candidate gene for motor neurone disease.

- Savage, A.L., Wilm, T.P., Khursheed, K., Shatunov, A., Morrison, K.E., Shaw, P.J., Shaw, C.E., Smith, B., Breen, G., Al-Chalabi, A. and Moss, D.J., 2014. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS.

Contents

List of Abbreviations

ALS	Amyotrophic lateral sclerosis
AMP	Ampicillin
AP	Anterior posterior axis
ATP	Adenosine-5'-triphosphate
Bp	base pair
β-ME	β-mercaptoethanol
BLAST	Basic Local Alignment Search Tool BLAST
BMPs	Bone morphogenetic proteins
C9orf72	Chromosome 9 open reading frame 72.
CNS	Central Nervous System
DAPI	4,6-diamidino-2-phenylindol
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dTTP	Deoxythymidine triphosphate
E	Embryonic day, number of days following start of incubation
ECR	Evolutionary Conserved region
ESTs	Expressed sequence tags
EDTA	Ethylenediamineteraacetic acid
FALS	Familial amyotrophic lateral sclerosis
FTLD	Frontotemporal lobar degeneration
FUS	*Fus*ed in *S*arcoma/Translocated in Sarcoma
GFP	Green Fluorescent protein
GWAS	Genomic-wide association study
HERV	Human endogenous retrovirus
HSP	Hereditary spastic paraplegias
Hb9	Homo box –Hb9 (Note: Official name MNX1)
Hb9 p1229	Hb9 enhancer sequence and p1229 plasmid construct
Hox	Home box
Ifs	Intermediate filaments
IRES	Internal ribosome entry site
INT1	Intron 1
Kb	Kilo bases
KIFAP3	Kinesin Associated Protein 3
LB	Lysogeny broth
LINE	Long interspersed element
LMN	Lower motor neurone
LTR	Long terminal repeats

MCS	Multiple Cloning Site
M	Molar
MND	Motor neurone disease
MAP	Microtubule-associated proteins
MAPT	Microtubule-associated proteins Tau
MNs	Motor neurons
miRNA	micro RNA
MNX1	Motor neurone and pancreas homeobox 1
mRNA	Messenger RNA
MTSS	Major Transcription Start Site
NaCl	Sodium Chloride
O D	Optical Density
OCT medium	Optimal cutting Temperature medium
ORF	Open reading frame
PBP	Progressive bulbar palsy
PBS	Phosphate buffered saline
PCR	Polymerase chain –reaction
PD	Parkinson's disease
PMA	Progressive muscular atrophy
p1229	A non-commercial reporter vector based on a p Bluescript backbone
PP	Processed pseudogene
PP	Proximal promoter
RA	Retinoic acid
RE	Restriction enzyme
RNP	Ribonucleoprotein complex
SAP	Shrimp alkaline phosphatase
Shh	Sonic hedgehog
SINE	Short interspersed element
SNP	Single nucleotide polymorphism
SOD1	Superoxide Dismutase 1
SALS	Sporadic amyotrophic lateral sclerosis
SVA	SINE- VNTR Alu
T_A	Annealing temperature
TARDBP	TAR DNA binding protein
TAE	Tris acetate EDTA buffer
TE	Transposable element
TF	Transcription factor
TPRT	Target primed reverse transcription
TR	Tandem repeat

TSS	Transcription Start Site
UMN	Upper Motor Neurone
UTR	Untranslated region
VNTR	Variable number tandem repeat

List of Figures

List of Tables

Abstract

Motor neurone disease refers to a group of neurological disorders that result from progressive degeneration of the motor neurones leading to death from respiratory failure within 3-5 years from the onset of symptoms. Amyotrophic lateral sclerosis (ALS) is one of the most common forms of motor neurone disease, and can be described as entailing the involvement of both upper and lower motor neurones. Usually, the disease is obvious and presents as asymmetrical weakness in the limbs and progressive muscular atrophy.

Over the last two decades more than 30 genes have been identified as involved in ALS pathology. These include PARK7 and FUS (Fused in sarcoma) which are the subject of this thesis. FUS is a multifunctional protein that has ubiquitous expression and is involved in different steps of RNA processing such as mRNA and miRNA. 10% of ALS cases are heritable and mutation of the FUS gene is found in 3-5% of familial forms of ALS. Therefore the FUS gene is important for it association as a candidate gene that is postulated to be important for ALS, in addition to various types of cancer. Mutations with FUS gene have been found in autopsy samples from the brain and spinal cord of patients with ALS. Understanding the regulation of FUS gene expression may, therefore, give insights into how its stimulus inducible expression may be associated with neurological disorders.

My hypothesis is that the evolutionary conserved regions (ECRs) and primate specific retrotransposons of the SINE-VNTR Alu (SVA) are regulatory domains from the human FUS gene and PARK7 genes. Consequently, the aim of this thesis was to develop an *in vivo* model to validate their regulation. Comparative genomic analysis was used between distant species, utilizing the ECR browser and UCSC browser to identify conserved regions from the FUS gene. It is demonstrated that ECR and SVA,

which can drive reporter gene activity *in vitro* (in the neuroblastoma cell line), are also capable of driving reporter gene expression in the chick embryo neural tube and brain at embryonic day 5.

In conclusion, these identified important ECRs from human FUS gene act as regulatory domains. In addition, the SVAs, representing the most recent retrotransposon to enter the human genome also was showed to have regulatory properties in FUS and PARK7 genes. Furthermore, the thesis demonstrates that these elements regulate gene expression *in vitro* and *in vivo*. This supports the idea that employment of the chick embryo as a useful and informative model system to analyse mammalian gene regulation.

Chapter 1

General Introduction

1.1 Overview

To understand nerve cell gene regulation, a description of both the embryonic development of the nervous system and adult nervous system is required. In the first part of the introduction, I describe the development of the human nervous system from the ectoderm. After that, the definition of Motor Neurone disease and classification of motor neurone disease is clarified. Next, a commentary on the genes implicated in motor neurone disease, including FUS and PARK7 and how they are implicated in ALS and neurodegenerative disease. I have focused specifically, on the elements that regulate gene expression. These regulatory elements are evolutionary conserved regions (ECRs) and the SVA. Finally, a brief overview of the model system used is provided.

1.2 Differentiation of neurons in the central nervous system

At the early embryonic stages, the human nervous system is developed from the external germinal layer of the blastula called the Ectoderm (Strominger et al., 2012). The ectoderm is induced by the trophic factors derived from the mesoderm and notochord to differentiate to neuroectoderm and form neural tissue. The neural tissue in the midback region will differentiate and thicken to form the neural plate which itself will develop into the central nervous system (Yuan et al., 2006). The neural plate will elongate and enlarge at the anterior region and form the brain, while the lateral edge of the neural plate will develop and enlarge to form the neural tube. The establishment of the neural plate is governed by the expression of four types of proteins that bind to the bone morphogenetic protein 4 (BMP-4): noggin, follistatin, chordin and bone morphogenetic protein (BMP) (Strominger et al., 2012). Molecular genetic studies of the development and establishment of the nervous system have revealed the role of genes in initiating the patterns and regions of the nervous system (Hatten, 1999). The differentiation of the neural tube and central nervous system (CNS) are under the control of a number of genomic transcription factors. Indeed, these factors within the regional differences in the underlying mesodermal somites have a role in the initiation of the anterior posterior axis (AP) and the formation of the forebrain, midbrain, hindbrain and spinal cord, while the development of the pattern of the dorsal-ventral axis is induced by different transcription factors. These factors will divide the neural tube into the dorsal and basal plates, for example the sensory neurones are produced from the dorsal plate while the motor neurones are produced from the basal plate. The development of lower motor neurons within the neural tube is influenced by factors from the notochord. These factors, including Sonic hedgehog

(Shh), fibroblast growth factor (FGF) and retinoic acid, play an important role in the process of dorsal and ventral neural tube specification. Shh originates and is synthesized in the notochord and can function in the ventral specification of the neural tube. In contrast, FGF originates from the dorsal ectoderm. Both of these factors can induce a second signal centre located within the neural tube (Pierani et al., 1999). Initially, Shh can influence the medial hinge cells to form the floor plate of the neural tube (see Figure 1.1)(Placzek et al., 1990). In turn these floor plate cells can also secrete Shh, with the most ventral part of the neural tube having the highest concentration of Shh (Liem et al., 1995, Sander et al., 2000). Cells that are located next to the floor plate are exposed to the highest concentration of Sonic hedgehog and will develop to form the ventral (V3) neurons and the second set of cells, which receive a lower concentration of Sonic hedgehog protein, concentration will develop into motor neurones. The other two sets of cells, meanwhile, which are exposed to even less Shh, will develop into V2 and V1 interneurons (see Figure 1.2) (Ericson et al., 1997).

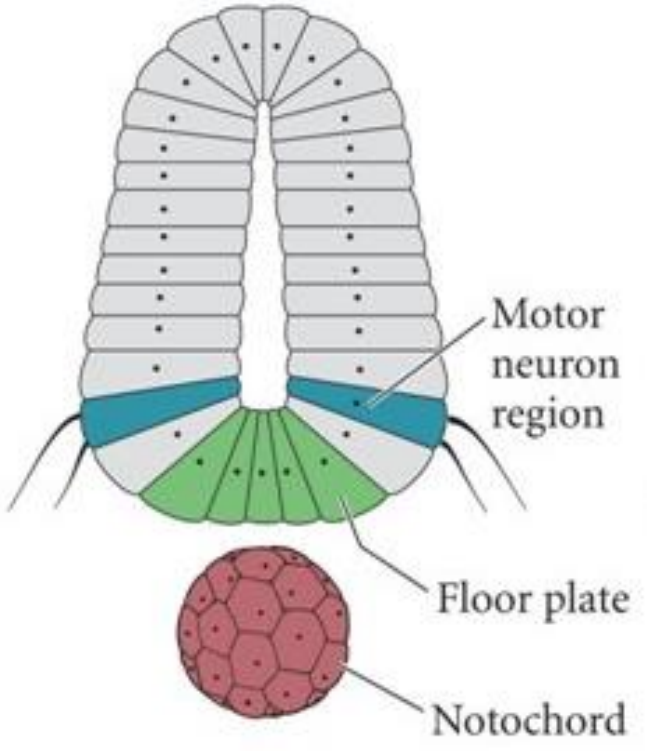


Figure 1.1: A newly formed vertebrate neural tube with two cell types of chick embryo neural tube. Those cells located near to the notochord develop into ventral floor plate and those in the ventrolateral sides are called the motor neurone region. Figure has been adapted from Placzek et al. (1990).

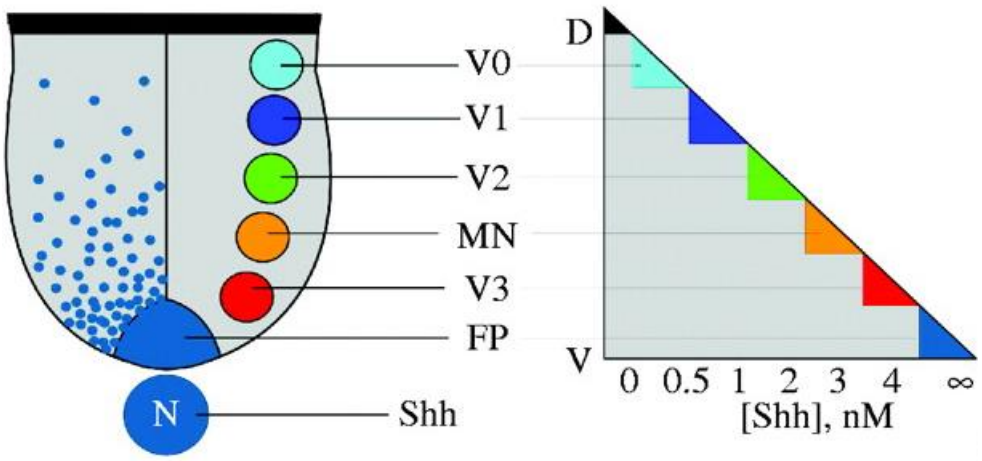


Figure 1.2: *In vitro* experimental illustration of the importance of Sonic hedgehog in the patterning of the ventral portion of the vertebrate neural tube. The relationship between the Sonic hedgehog gradient concentration, the development of specific neuronal types *in vitro*, and the space in nM from the notochord. This Figure has been adapted from; Placzek et al. (1990), Ericson et al. (1997).

The basic organization of the neural tube into a brain, meanwhile, is a result of the expression of particular genes encoding protein factors that regulate the transcriptions of other genes and control brain differentiation. HOX genes are one such example. These control the hind brain cell identity and initiate the segmental pattern of the spinal cord and spinal ganglia. HOX gene expression is regulated by retinoic acid (RA) and has an important role in the establishment of the AP axis in both the spinal cord and the hind brain. Another example is *Otx2* transcription factors, which are expressed in the anterior region of the forebrain to the border between midbrain and hindbrain. wnt1, meanwhile, is in control of the AP axis of these regions, the Pax6 gene is expressed in rostral (motor) regions of the cortex, and the Emx2 gene is expressed in the posterior part of the motor region of the developing cortex.

1.3 Classification of neurones

Neurones can be classified into sensory neurones, motor neurones and association neurones (Kaye and Davis, 2014). Sensory neurones have a role in transmitting the impulses to the central nervous system and can be in two subtypes; somatic sensory, which transmit the impulses from sensor receptors in the skin, bones and joint muscles to the central nervous system (CNS); and visceral sensory, which transmit the impulses from visceral organs. Motor neurones, however, transmit impulses away from the central nervous system and are divided into two subtypes, somatic motor neurons, which innervate the skeletal muscles, and visceral motor (autonomic motors), which innervate cardiac and smooth muscles and glands. The third subtype, associated neurones, or interneurones, transmit information, directly and indirectly, from sensory neurones to motor neurones and it is this type that account for most neurones in the brain (Stifani, 2014).

1.4 Motor neurones (MNs)

Motor neurones can be defined as a group of neuronal cells located in the central nervous system (CNS). These cells can regulate both voluntary and involuntary movements (Kaye and Davis, 2014). Motor neurones (MNs), generally, can be classified into two main categories according to the location of their cell body: upper motor neurones and lower motor neurones (Stifani, 2014). Despite the fact that both types share the same nomenclature, upper motor neurones and lower motor neurones are fundamentally different in terms of the location of their cell body, neurotransmitter, lesion symptoms and cell targeting (see Table 1.1) and finally clinical symptoms upon lesion (Talbot and Marsden, 2008). Furthermore, upper MNs

activate lower motor neurons, while lower MNs (except for visceral MNs) are connected to the target muscles directly.

Table 1.1: Comparison between the upper and lower motor neurones regarding, location, symptoms upon lesion neurotransmitters and targeting (Redrawn from (Stifani, 2014)).

	Upper MNs	**Lower MNs**
Location	Cortex and Brainstem	Brainstem and spinal cord
Neurotransmitter	Glutamate	Acetylcholine
Targeting	Within the CNS	Outside the CNS
Symptoms upon lesion	Spasticity	Paralysis

1.4.1 Upper motor neurones

Upper motor neurones originate from either the motor region of the cerebral cortex or from the brain stem (Figure 1.3) (Purves et al., 2001, Kenneth, 2003). Upper motor neurones can transmit motor information to the downstream target (lower motor neurones) through glutamatergic neurotransmitters. They are mainly initiators of voluntary movement, and reside in the fifth layer of the grey matter primary motor cortex, where they are called Betz cells. Betz cell or (pyramidal cells of Betz) are a type of giant pyramidal cell that have the largest cell bodies in the brain and are approximately 0.1 mm in diameter. In fact the Betz cells are located precisely in the precentral gyrus. The precentral gyrus is considered the most important region of the frontal lobe. Upper motor neurones project their axons away from the precentral gyrus and end in the brainstem. In the brainstem, they form synapses with the lower medulla oblongata and form the lateral corticospinal tract that is located on both sides of the spinal cord. Those fibres that do not form synapses, will pass into the corticospinal tract to form the anterior corticospinal tract (Purves et al., 2001, Kenneth, 2003).

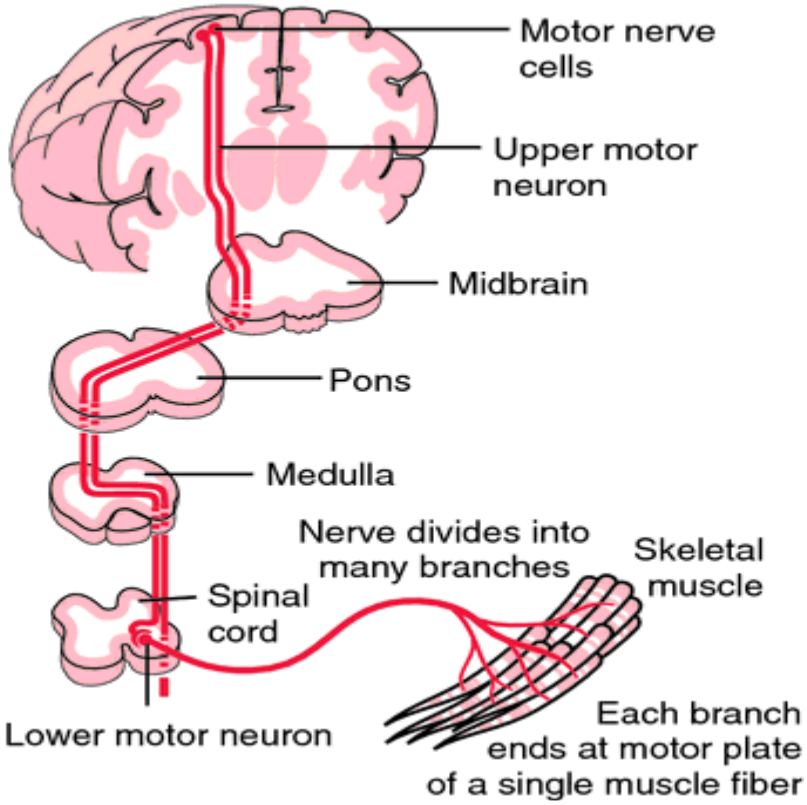


Figure 1.3: Illustration of the location of upper and lower motor neurones according to the anatomical pathways controlling movement. Upper motor neurones originate from the motor region of the cerebral cortex; they are located in the motor cortex and are responsible for voluntary movements. These neurones contribute their axons to the corticospinal tract. Then, the corticospinal tract passes the through the medulla (brain stem). After that, most fibres decussate and enter the lateral corticospinal tract and the remaining fibres that not cross will enter the ventral corticospinal tract directly. Then, the axons extend to reach their target in the spinal cord. These are the lower motor neurones, whose bodies are located in the grey matter and whose axons innervate the muscle fibres directly. This Figure has been adapted from; Damjanov (2000).

1.4.2 Lower motor neurones

Lower motor neurones cell bodies are located in specific nuclei in the brainstem and ventral horn of the spinal cord. Similar to the upper motor neurones, they reside within the CNS but, in contrast to upper motor neurones, they project their axonal extension and connection outside of the nervous system (Figure 1.3). Lower motor neurones receive neurotransmitter inputs from upper motor neurones, sensory neurones and interneurons. Furthermore, injury to lower motor neurones can cause paralysis since the damaged pathway cannot be replaced by other pathways to transport the

information to the downstream target in the periphery (Stifani, 2014). Lower motor neurones can be classified into three categories: I) Somatic motor neurones, ii) Visceral motor neurones and iii) Special visceral motor neurones (Kandel et al., 2000). Somatic motor neurones are located in the midbrain, hindbrain and ventral horn of the spinal cord. Somatic motor neurones innervate and control eye and tongue muscle movement and innervate limb muscles. The visceral motor neurones, or general visceral motor neurones, are located in the hindbrain and innervate the parasympathetic neurones that in turn innervate the tear and sweat glands, pulmonary glands, the cardiovascular and gastrointestinal system and, finally, the smooth muscles. Branchial motor neurones, or special visceral motor neurones, are located in the hindbrain and innervate the skeletal muscles that derive from the branchial arch muscles. They control the movement of jaw and facial expression, larynx and pharynx movements.

1.5 Motor neurones rely on motor proteins

When axons are viewed by quick-freeze-deep-etch electron microscopy their cytoskeletal structure (ultra-structural inclusions) shows three different types of filaments i.e. actin microfilaments, microtubules and intermediate filaments (IFs). The intermediate filaments, or neurofilaments, are abundant in many axons. The axonal cytoskeleton provides a mechanical route to transport the secreted molecules (in vesicles) produced in the cell body which is necessary for the synapse transmission in the axonal terminal (Yuan et al., 2006).

The highly specialised morphology of these neurones depends on a dynamic system that can transport the vesicles' protein and organelles, including mitochondria and secretory vesicles, from the cell body towards the axonal terminal by a mechanism called fast axonal transport. Slow axonal transport, meanwhile, can convey other cytoskeletal proteins towards the axonal terminal (Dillman III et al., 1996, Hirokawa and Takemura, 2004). The inner environment of the neurones has a machinery system which can be characterised by two main elements: the microtubules, which act as tracks and the molecular motors that can transport cargo along the microtubules. The microtubules have two ends, one called the plus end that is localised towards the synapse, while the minus end is towards the cell body. In the cell cytoplasm, dynein is considered to be a major motor protein, and is a retrograde transport that moves towards the minus end of the microtubules. A member of the kinesin motor family, meanwhile, moves the cargo towards the plus end of the microtubules (Rao et al., 2011). Microtubule motor proteins have an essential role in neurones and other cell types (Figure 1.4). There are several types of microtubule binding protein, including: kinesin and the kinesin superfamily, dyneins, dynactin and tau protein.

1.5.1. Kinesin superfamily proteins (KIFs) and Kinesin

Molecular motors have an essential role in transportation of the protein complexes, organelles and mRNA to a particular location. This family have more than 45 members known as kinesin superfamily proteins (KIFs), all these members have been identified in the human genome (Xia et al., 2003) and share homology with the 340 amino acid motor domain of conventional kinesin (Miki et al., 2001). KIFs are a microtubule motor protein and have an essential role in the anterograde transport in axons and dendrites. Kinesin can use the energy from ATP hydrolysis to convey the protein complexes, organelles and mRNA along the microtubules to target destinations (Goldstein and Yang, 2000). A mutation in the kinesin gene expressed in neurons can cause hereditary spastic paraplegia type 10 (SPG10) (Reid et al., 2002). Similar mutations were observed in other kinesin superfamily proteins, which also caused defects in motor activity, and resulted in axonal degeneration of motor and sensory neurons in patients with these mutations.

1.5.2 Dyneins

Cytoplasmic dynein in neurones has a major role in retrograde axonal transport but also has a variety of other cellular functions including neurofilament transport, mRNA localisation and endoplasmic reticulum-to-Golgi vesicular transport. Dyneins use the energy from ATP hydrolysis in order to move their cargo towards microtubule minus ends (Holzbaur, 2004, Yuan et al., 2006). Cellular studies have demonstrated that mutations in the cytoplasmic dynein heavy chain can cause progressive motor neuron degeneration in heterozygous and homozygous mutant mice and perturb neuronal function of dynein. These mutations can lead to inclusion of the Lewy-like body, which represents a key feature in human pathology (Hafezparast et al., 2003).

1.5.3 Dynactin

Dynactin may also be called dynein-activator complex and has a major role in the cellular function of cytoplasmic dynein. Dynactin can act to raise the efficacy of the motor and to pair it with its vesicular cargos (Holzbaur, 2004). Work by Puls et al (Puls et al., 2003) demonstrated that a substitution mutation in a single base pair can cause misfolding of dynactin's binding domain and lead to human motor neuron disease.

1.5.4 Tau protein

Tau was identified in 1972 for its function as a promoting factor in tubulin polymerisation (Weingarten et al., 1975). Tau is a neuronal protein and is located primarily in axons. Tau has an important role in stabilising and organising the dynamics of microtubules (Rao et al., 2011). The pathophysiology of tau protein was studied by Orozco et al (Orozco et al., 2012) which showed that tau transcripts including exon 10 can cause frontotemporal dementia and Parkinsonism.

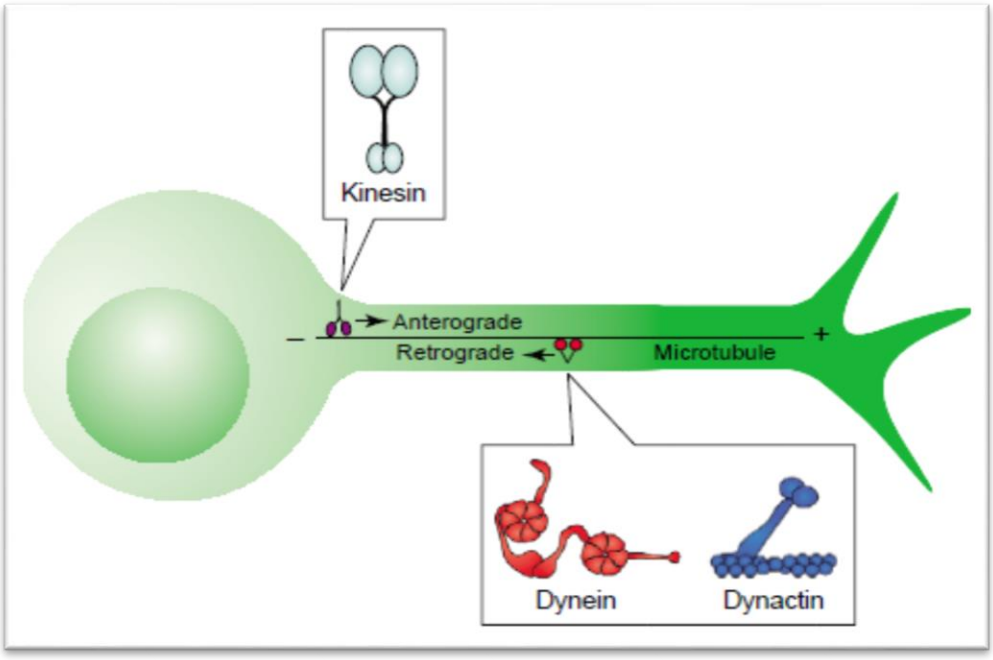


Figure 1.4: Microtubule motor proteins shuttling the cargo in the neuronal axon. In neurones, the microtubule motor protein can transport the material toward the axon presynaptic terminal by anterograde transport powered by kinesin and proteins of the kinesin super family (KIF). Retrograde axonal transport requires both cytoplasmic dynein and dynactin in order to transport the material back to the cell body for degradation. This Figure has been adapted from; Holzbaur (2004).

1.6 Motor Neurone Diseases (MNDs)

Motor neurone disease (MND) is a progressive neurodegenerative disease, ultimately leading to death with respiratory failure, commonly five years from the onset of the symptoms (Aoun et al., 2012) . The terminology MND is typically used in the UK and represents an umbrella that collects all types of motor neurone disease, while the term Amyotrophic Lateral Sclerosis or Lou Gehrig's disease is commonly used in the USA. For this reason MND and ALS are often used interchangeably although in this thesis MND is used as an umbrella term and ALS as a specific type, as defined here. The symptoms associated with MND can be characterised by motor weakness and bulbar dysfunction (Bäumer et al., 2014).

The existence of a disease associated with degeneration in the motor neurones was first observed in the 19th century, and the neurologist Jean-Martin Charcot was the first person to identify the combination between upper and lower motor neurone dysfunction (Vejjajiva et al., 1967). The incidence of motor neurone degeneration in the UK population is about 1 in 50 000 (Bäumer et al., 2014), with currently between 5000 and 6000 people living with diagnosed MND in the UK. MND is therefore considered to be a relatively rare disease (Fisher et al., 2012). MND is also an age-related disease, with most affected individuals being between 50 and 65 years old; however, 5-10% of patients are below 40 years (Bäumer et al., 2014) . Men are usually more susceptible to MND (one and a half times more so than women). MND is caused by a combination of genetic and environmental factors and many theories have tried to explain the role of dietary factors in MND. For example lathyrism in India and konzo in Africa are examples of neurotoxic disorders mainly caused by the intake of the neurotoxic plant products *Lathyrus sativus* and *Manihot esculenta* (Spencer and

Palmer, 2012). These toxins can affect the Betz cells specifically. Environmental poisons such as pesticides and heavy metals, in combination with genes that have risk factors for disease susceptibly may also be involved (Trojsi et al., 2013) .

1.7 Classification of motor neurone disease

In general, MNDs are considered to be a syndrome that involves various clinical phenotypes, which share features of progressive motor neurone degeneration. Many criteria are used to classify MNDs including: degree of involvement of upper motor neurones (UMN) or lower motor neurones (LMN). It can be classified into sub forms for example: a) Amyotrophic lateral sclerosis (ALS), b) Progressive bulbar palsy (PBP), c) Progressive muscular atrophy (PMA), d) Primary lateral sclerosis. e) Pseudobulbar palsy, f) Hereditary spastic paraplegias. Other criteria that can be used to classify MNDs include: rapidity of the disease progression, e.g. slow or rapid. Diffuse MND or regional spread e.g. flail arm, bulbar (PBP). Age at onset: younger than 45 years or older, familial or sporadic, cognitive involvement: Frontotemporal dementia (FTD) or loss of executive function (Ludolph et al., 2015) . There is no precise test for MND and it is known that MND has variable characteristics that might affect people in different ways. These include the distribution of the disease in the body and also the rate of disease progression (Ludolph et al., 2015). Clinical diagnosis is the only way to classify motor neurone disease.

1.7.1 Amyotrophic lateral sclerosis (ALS)

ALS is the typical or classical form of MND. It can be described as having involvement of both upper and lower motor neurones (Figure 1.5). Initially, symptoms start in the hands and feet as asymmetrical weakness, then the muscles tend to become stiff and weak, followed by wasting in the limbs due to damage in the motor neurones located in the brain, brainstem and corticospinal tract for instance. It can also present as spinal onset with limb weakness or bulbar onset with difficulty with speech and swallowing (Kelly, 2012, Xia et al., 2003, Agosta et al., 2015) finally leading to death from respiratory failure.

1.7.2 Progressive bulbar palsy (PBP)

This form of MND involves both upper and lower motor neurones (Cerero Lapiedra et al., 2002). At first, the motor neurone cells that are located in the motor cortex, spinal cord, brain stem and pyramidal tracts degenerate (Cerero Lapiedra et al., 2002). Then, it affects the bulbar muscles, including the muscles used for talking, chewing and swallowing. Patients with this form of the disease exhibit bulbar and cortico-bulbar symptoms, which include dysphagia, dysarthria, tone wasting and brisk jaw jerk.

1.7.3 Progressive muscular atrophy (PMA)

This is an uncommon form of MND, about ~ 4 % of people with MND have this type. This condition is also known as lower motor neurone syndrome. The muscles of the hands and feet are usually first affected or it might affect just one of either. This form is applied when there is no involvement of upper motor disease and usually has a better prognosis (Riku et al., 2014).

1.7.4 Primary lateral sclerosis

This is a rare type of MND and accounts for 1% of all cases. It can mainly be diagnosed as weakness and spasticity in the leg muscles and spastic bulbar weakness (Figure 1.5). Typically, symptoms start in one leg and then diffuse over time to involve the arms, speech and swallowing muscles (Turner et al., 2013). In fact, in most types of MND, the symptoms start in a certain location and spread out as they progressively affect neighbouring neuronal cells (Talbot, 2002).

1.7.5 Pseudobulbar palsy

This condition is characterised mainly by failure to control facial muscles, specifically the muscles which control the movements needed for chewing and speaking. People affected with this disorder can show difficulty in chewing and swallowing. This condition results from degeneration of motor neurones located in the brain stem and corticobulbar tract (Steele et al., 2014).

1.7.6 Hereditary spastic paraplegias (HSP)

This is a group of conditions mainly characterised by pyramidal weakness and presence of spasticity in the lower limbs and hyper-reflexia with extensor plantar responses (Fink, 2013). HSP has been found in all types of inheritance (autosomal dominant, autosomal recessive and X-linked recessive (Fink, 2013)

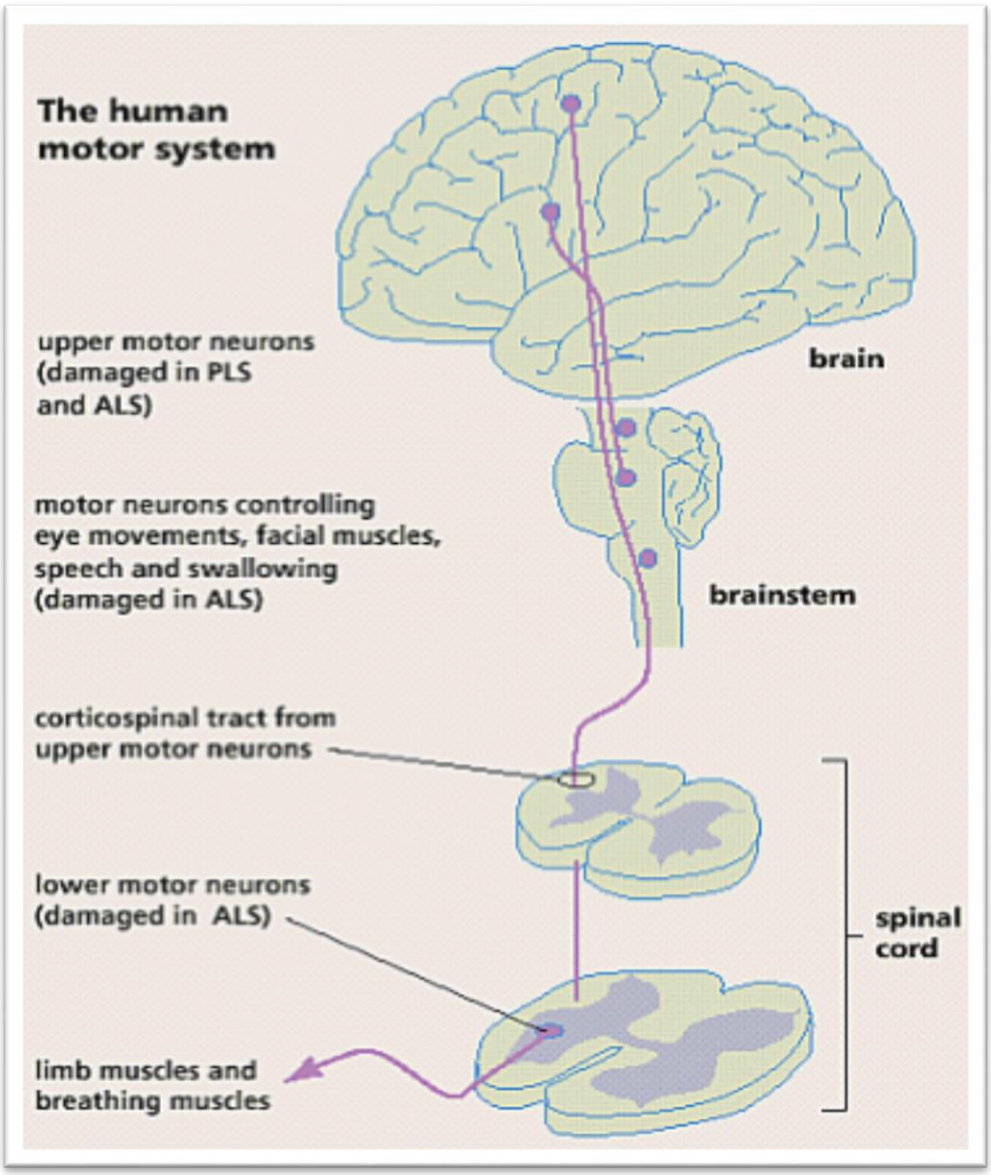


Figure 1.5: Illustration of the location of motor neurone cells in the nervous system with relation to the MND form. The upper motor neurones which are located in the motor cortex. In the case of ALS and PLS these cells will be affected and eventually damaged. The fibres descending from the upper motor neurones will pass through the corticospinal tract and synapse with lower motor neurones; in ALS cases these neurones are also affected and eventually become damaged. This Figure has been adapted from; Stifani (2014).

1.8 Genes implicated in ALS.

Over the last two decades more than 30 genes have been identified and characterised for their role in ALS pathology. This section will explore several mutant genes that have been implicated in familial amyotrophic lateral sclerosis (FALS). These include Superoxide Dismutase 1 **(**SOD1), TAR DNA binding protein (TARDBP), Elongation Protein 3 (ELP3), Kinesin-associated protein 3 (KIFAP3), chromosome 9 open reading frame 72 (C9orf72) and Fused in sarcoma (FUS). Along with these genes, the PARK 7 gene was also studied due to its association with both motor neurone disease and neurodegenerative diseases more generally.

1.8.1 Superoxide Dismutase 1 (SOD 1).

Mutations in the SOD1 gene were first identified in relation to ALS in 1993 (Rosen et al., 1993). They are implicated in 20% of autosomal dominant cases of FALS and are also found in all types of motor neurone disease (Dion et al., 2009). In addition, a few cases of ALS related to SOD1 mutation show a recessive inheritance (Dion et al., 2009). SOD1 is an abundant and ubiquitously expressed gene, encoding for the copper/zinc superoxide dismutase enzyme that functions as an active homodimer cytosolic enzyme. The main role of this enzyme is to reduce the production of reactive oxygen species (ROS) by catalysing the conversion of toxic superoxide radicals to molecular oxygen and hydrogen peroxide. The protein consists of 153 amino acids and changes in more than 125 distinct amino acids have been linked to ALS cases. Mutations within this gene might disrupt the highly-stable native SOD1 enzyme, leading to protein misfolding and then aggregation. Generally, oxidation and mutation of SOD1 can lead to misfolding of the SOD1 protein producing toxic gain-of function (Grad and Cashman, 2014). As a result of SOD1 misfolded protein, SOD1 has been

implicated in ALS pathology in the form of excitotoxicity, mitochondrial dysfunction, endosomal trafficking, oxidative stress and axonopathy (Beckman et al., 1993, Boxer et al., 2011).

1.8.2 Transactive response DNA binding protein (TARDBP coding for TDP-43)

The transactive response DNA binding protein (TDP-43) is encoded by the TARDBP gene. This protein has ubiquitous expression and is a 43 kD nuclear protein that translocates between the nucleus and the cytoplasm. TDP43 plays an important role in gene regulation, including in transcription, RNA splicing, mRNA stability and micro RNA biogenesis (Dion et al., 2009). Deletion of TARDPB results in embryonic lethality and loss of function of TDP-43 in neuronal cell culture, causing mRNA missplicing and neuronal defects (Mutihac et al., 2015). Mutations in the gene coding for TDP-43 have been identified as having a significant causative link for ALS and as accounting for 4% of FALS and some cases of FTD, as well as involvement in all types of motor neurone disease (Kabashi et al., 2008). TDP-43 consists of 414 amino acids, is encoded by six exons and has two RNA recognition motifs and a C–terminal that is enriched with a glycine region and that interacts with heterogeneous nuclear ribonucleoproteins (hnRNPs) (Buratti et al., 2005). Dominant mutations in the TARDBP gene have been found in ALS cases (Lagier-Tourenne and Cleveland, 2009, Gitcho et al., 2008, Kabashi et al., 2008, Sreedharan et al., 2008, Van Deerlin et al., 2008). These studies have demonstrated that the aberrant form of TDP-43 can cause neurodegeneration. Most mutations are located in the C-terminal which is encoded by exon six of the TARDBP gene. Patients with mutation in TDP-43 develop ALS, and over 50% also developed cognitive impairment (Colombrita et al., 2009).

Furthermore, TDP-43 inclusions were found in the cytoplasm or nucleus of neurone and glial cells from the spinal cords and brains of ALS patients (Neumann et al., 2006, Van Deerlin et al., 2008). Neuropathy caused by TARDBP mutations can be characterized by hyper-phosphorylation and ubiquitination with the generation of a 25 kD C-terminal fragment that is missing the nuclear localised domain. In 2008, dominant mutations in the TARDBP gene were identified as a major cause of ALS and represent approximately 3% of familial ALS cases and 1.5% of sporadic cases (Dion et al., 2009). Inclusion bodies associated with mutation in the TDP-43 are seen in the cytoplasm of both the upper and lower motor neurones and both the frontal and temporal cortex of the central nervous system (Al-Chalabi et al., 2012).

1.8.3 Elongation protein 3

ELP3 gene codes for an RNA processing protein(Simpson et al., 2009). It has an essential role in the construction of the Elongator complex, called holo-Elongator, and is a component associated with RNA polymerase II, which is essential for RNA binding (Greenwood et al., 2009, Kristjuhan and Svejstrup, 2004, Tanner and Linder, 2001). Two loss-of-function mutations were identified in Drosophila ELP3 which can cause abnormality in axonal targeting and synaptic development. Knockout of ELP3 in zebra fish resulted in a shortening the motor neurone branching and other abnormal features of the nervous system (Simpson et al., 2008). Added to that in an association study of 1483 individuals, ELP3 was found to have an association with human motor neurone degeneration in the form of ALS in three different populations (Simpson et al., 2008).

1.8.4 Kinesin Associated Protein 3

KIFAP3 gene codes for kinesin associated protein. This gene, located in the chromosome1 q24.2, is involved in chromosome movement and has an intracellular function as part of the transport machinery. Genomic wide association studies by (Landers et al., 2009) have demonstrated that the CC genotype of the single nucleotide polymorphisms (SNP) rs1541160 located within intron 8 of the KIFAP3 gene is associated with a survival advantage of 18 months among 1014 patients with ALS. Further analysis of this study showed that there is a linkage disequilibrium between rs1541160 and rs522444 in KIFAP3 promoter region and these alleles are favourable in correlation with the decreased level of KIFAP3 expression in brain tissue and lymphoblast of ALS patients.

1.8.5 Chromosome 9 open reading frame 72

The gene termed Chromosome 9 open reading frame 72 (C9orf72) was discovered by (Gijselinck et al., 2012) through its linkage to ALS, which is achieved through a large hexa-nucleotide coded for GGGGCC repeat expansion mutations residing in the non-coding region of the C9orf72 gene that has a correlation with neurodegenerative disease including fronto-temporal lobar degeneration (FTLD) and ALS. Mutations in C9orf72 account for approximately 20-30% of FALS patients. The pathology associated with this gene can result from the loss of function of the C9orf72 protein. Alternatively, it can be caused by gain of function mechanisms including sense or antisense repeats of RNA (Mizielinska et al., 2017). Another mechanism may be the formation of RNA aggregates in the neuronal nuclei which might sequester important RNA-binding proteins and form a toxic dipeptide repeat (DPR) protein in order to

mediate the RNA repeat translated with the absence of an ATG initiation codon (Gijselinck et al., 2012, Van Mossevelde et al., 2017).

1.8.6 PARK7

The PARK7 gene is located on the distal part of the short arm of the chromosome 1p36.2-p36.3 and the gene is also called DJ-1. It was first identified as a protein that has a potential oncogenetic role by working in cooperation with *ras* oncogene to transform mouse NIH-3T3 cells (Nagakubo et al., 1997). PARK7 protein is ubiquitously expressed in over 22 human tissues and is also found to be expressed in both cell nuclei and the cytoplasm of the HeLa cell line (Nagakubo et al., 1997, Singh et al., 2015). PARK7 is highly expressed in brain and cerebral tissues. The product of this gene has 189 amino acids. The protein can act as a positive regulator of androgen receptor-dependent transcription AR, and thus it can interfere with the binding of the PIASxα with AR (Takahashi et al., 2001).

PARK7 has an important role in the human cell lines, as an antioxidant and or a molecular chaperone. Many studies have suggested that PARK7 may function as a redox-sensitive chaperone, and can protect the neurons from oxidative stress and cell death. Defects with the PARK 7 gene can cause several diseases, including Parkinson's disease 7, autosomal recessive early onset familial Parkinson's disease (PD), juvenile type 2se (Dawson and Dawson, 2003, Alvarez-Castelao et al., 2012). Furthermore, PARK7 can cause breast cancer and non-small cell lung carcinoma (Le Naour et al., 2001, MacKeigan et al., 2003b). PD is known as a heterogeneous disease, with most cases appearing to have sporadic origins, although mutations with specific genes can cause familial PD (Macedo et al., 2003). PD symptoms can have general

features, such as resting tremor, postural instability with muscular rigidity, and akinesia (Jankovic, 2008).

1.8.7 Fused in Sarcoma Protein (FUS).

The FUS gene was selected to be the focus of the investigation in this thesis for a number of reasons. At the beginning of this study, the FUS gene had recently been discovered to be implicated in ALS (Kwiatkowski et al., 2009) and it was therefore clearly an important next step to know the role of the FUS gene in ALS; secondly, FUS mutation had been shown to represent more than 3-5% of cases of FALS (Kwiatkowski et al., 2009). Thirdly, we identified an SVA with central VNTR located upstream of the FUS gene that might have an important role in the regulation of the FUS gene (Savage et al., 2014a).

The FUS/TLS gene was identified twice; first in human myxoid liposarcoma, where it was named FUS (Fused in sarcoma) since the gene was found to be disrupted in the chromosomal translocation, t(12;16) (q13;p11) and fused to CHOP gene (GADD153), which acts as a transcriptional repressor (Rabbitts et al., 1993) and later in human myeloid leukaemia, when the gene was fused with the transcriptional activator, ERG and seen in the recurrent chromosomal translocation, t(16;21) (p11:q22) . FUS gene coding for Fused in Sarcoma Protein is located on chromosome 16. FUS/TLS and $hTAF_{11}68$ are known as RNP-motif-containing proteins and are present in the Pol II transcription complexes and involved in Pol II transcription (Bertolotti et al., 1996). The collection of human FUS/TLS, with EWS protein and TAF 15 (TATA-binding protein associated factor) with *Drosophila Cabeza*/SARFH protein comprise a sub-family of RNA binding proteins, known as the TET or FET family (TLS/FUS, Ewing's sarcoma, TATA-binding protein-associated factorII-68) (Bertolotti et al., 1996).

Previous studies have determined that the FUS/TLS, EWS, RBP56/hTAFII68 genes have conserved domain structure proteins and similar genomic structures, which might indicate that these genes originated from the same ancestor gene (Morohoshi et al., 1998). FUS/TLS protein has been reported to form a ternary complex with other proteins such as hnRNP A1 and hnRNP C1/C2 (Zinszner et al., 1994). It is also worth noting that FUS/TLS protein was identified by (Calvio et al., 1995) as hnRNB P2 assembled on pre-mRNA. While FUS/TLS-like protein was isolated from HeLa cell as a TATA-binding protein associated factor (Khokha and Loots) by (Bertolotti et al., 1996). Added to that it was demonstrated by (Bertolotti et al., 1996) that EWS protein was correlated with TFIID and the RNA polymerase II complexes. *Drosophila Cabza/SARFH* protein also has a similar structure to EWS protein, containing active transcription units through RNA polymerase II (Immanuel et al., 1995). These data suggested that FUS/TLS, EWS and *Drosophila Cabeza/SARFH* with hTAFII68proteins are involved in the transcription and metabolism of mRNA. In the TET family, the RNA binding domain of most of its members is encoded by three exons; the exception is the *Cabeza/SARFH* protein where the RNA binding domain is encoded by two exons. The TET family, therefore, contains the only RNA binding domain that has a strongly conserved exon structure, and this conservation is not seen in other families of RNA binding proteins (Morohoshi et al., 1998).

The FUS/TLS gene consists of 15 exons that span 12 kb of human genomic DNA. The translation initiation codon is located in the first exon, the promoter sequence has no TATA box-like sequence but does have a GC rich region and multi binding site for transcription factors like Activator Protein (AP2), specificity protein 1 (Sp1) and GC-rich sequence DNA-binding factor (GCF) located in the promoter region (Åman et al.,

1996). The FUS/TLS consists of 526 amino acid proteins. Figure 1.6 illustrates the main characteristics of this gene, including an N-terminal domain that has four protein domains; a region SYQG, which consists of serine, tyrosine, glutamine and glycine and a RNA recognition motif (RRM); a multiple RGG region that is implicated in RNA binding; a C terminal repeat which consists of a zinc finger motif; and a highly conserved C-terminal region located at the very end (Figure 1.6) (Pérez-Mancera and Sánchez-Garía, 2004).

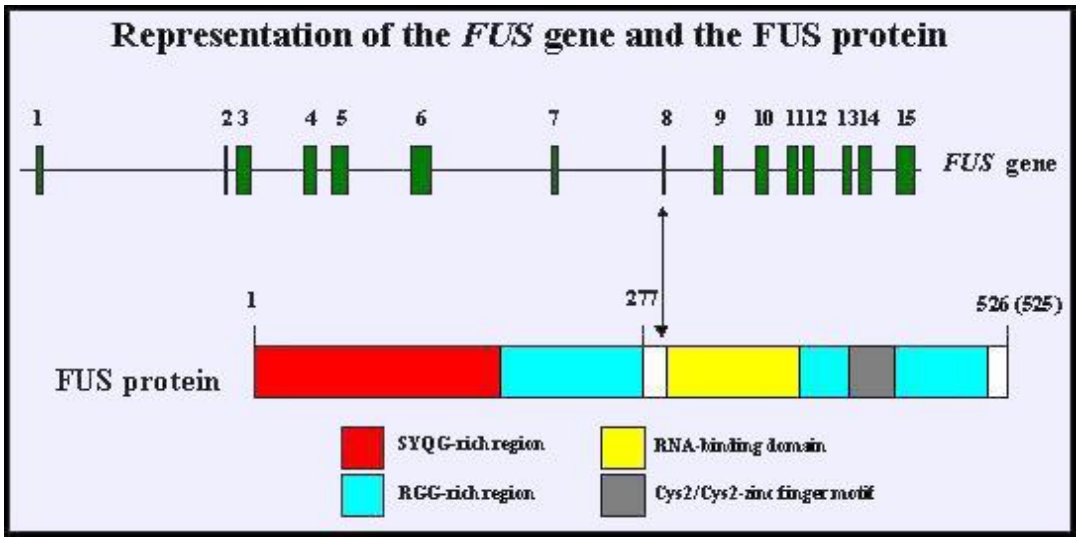


Figure 1.6: Schematic view showing the structure of the FUS gene together with the associated FUS protein structure. The FUS gene contains 15 exons, while the FUS protein comprises 526 amino acids. The FUS protein is including in an N-terminal domain that is enriched in four types of amino acid: serine, tyrosine, glutamine and glycine (the SYQG region), a glycine rich region, one RNA recognition motif (RRM), a multiple RGG-rich region that is implicated in RNA binding, a C-terminal repeat which consists of a zing finger motif and a highly conserved region located at the very end of the C-terminal region. This Figure has been adapted from; Perez-Mancera and Sanchez-Garcia (2004).

1.8.8 FUS mutational analysis

In a previous study conducted on multi-generational British kindred linked with a FALS study exons from the FUS gene were sequenced (Waibel et al., 2010). This

study identified a single base pair change in exon 15 of the FUS/TLS gene, with this change leading to a substitution of arginine to cysteine at location 521(R521C), a residue at the very C-terminal end of the protein, which is a highly conserved region, functioning in the area of nuclear localisation signals (NLS).

The vast majority of FUS mutations that have been identified in the FALS cases are missense mutations. There is also, however, a novel nonsense mutation (R495X) in the C-terminal and it has been predicted that this mutation may lead to production of truncated TLS/FUS and the deletion of the sequences that encode the NLS domain (Waibel et al., 2010). This results in TLS/FUS mislocalisation and the formation of the cytoplasmic inclusion which is a pathological hallmark of neurons that have been affected by FUS mutation. Most FUS/TLS mutations are dominant, however a rarer recessive mutation (H517Q) was found in a family of Cape Verde origin (Sproviero et al., 2012).

Investigation of the autopsy samples of brains and spinal cords from patients with FUS/TLS mutations has shown abnormal FUS/TLS to be present in the cytoplasm of neurons and glial cells, and lower motor neurones in the spinal cord (Sproviero et al., 2012, Urwin et al., 2010, Deng et al., 2010). Mutant human ALS FUS was found to be overexpressed in adult transgenic rats, causing a loss in neurones in both the cortex and hippocampus leading to learning and memory deficits (Huang et al., 2011).

A number of mutations residue in the 3'UTR of the FUS gene in ALS patients at a rate of 1.2%, as demonstrated by (Sabatelli et al., 2013b). These mutations increase the level of the FUS protein in a limited number of cases and since such protein over abundance has been correlated with the disease severity, it is proposed that the level of FUS protein has a crucial role in the cell phenotype and in the development of

neurological disorders, including ALS (Mitchell et al., 2013, Neumann et al., 2009, Mackenzie et al., 2010). Over expression of the wild type of the FUS gene in a transgenic mouse model has in fact been shown to cause degeneration in motor neurones in an age and dose-dependent fashion (Mitchell et al., 2013, Mackenzie et al., 2010).

1.8.9 Implication of FUS mutation for ALS and FTLD

ALS and FTLD are related neurodegenerative diseases. ALS can be considered degeneration of the upper and lower motor neurone cells, causing skeletal muscle weakness and limb paralysis, while FTLD affects cortical neurones leading to cortical dementia (Forman et al., 2006). ALS symptoms can overlap with FTLD pathology and, in some cases; ALS patients can develop sporadic FTLD. Both conditions also have ubiquitin-positive inclusion (Deng et al., 2010, Neumann et al., 2009, Urwin et al., 2010). In a cell line, FUS knockdown appears to cause preferential inclusion of tau exons 3 and 10 and result in disruption of tau's role. A similar study by (Orozco et al., 2012) has demonstrated that the inclusion of tau exon 10 can also lead to frontotemporal dementia and Parkinsonism (Hutton et al., 1998). Tau splicing is therefore a potential target of FUS function in neurones (Orozco et al., 2012). All these studies have demonstrated that there are several mechanisms that have a major role in controlling FUS expression; these mechanisms can be at the level of the FUS expression or the expression of isoforms, or might take place post transcription or translation, but these all have their role to play in neurological disorders. For example, the mutant form of FUS protein or the regulation of the gene transcription or translation process results in misplaced of the FUS protein. Result in accumulation in the cytoplasm rather than in the nucleus of the cell. This was happen due to FUS

mutants in a damaged or missing nuclear localization sequence (NLS) (Gal et al., 2011). The pathological accumulation of the mutant form of the FUS gene in stress granules or the formation of the globular and skein-like inclusions of the wild type of FUS gene can cause the death of motor neurones (Gal et al., 2011). For more than two decades researchers have focused on identifying the mutations and polymorphisms (Breen et al., 2008b) in the exonic sequences of proteins in order to define the functional associations with disease. Over 95% of these studies have demonstrated the relationship of the polymorphisms and mutations. Mutations can be defined as a polymorphism when they occur in at least 1% of the population affecting the coding regions of the associated genes with the risk of disease. Many mutations and polymorphisms that have a strong correlation with the progression of neurological disorder, however, can occur in genes in the regions of noncoding DNA sequence and even at some distances from the transcriptional start site. Most previous studies have focused on mutations in the FUS protein and relatively little is known about FUS gene regulation. Therefore, the regulation of the FUS and PARK7 genes were the main theme of this thesis. Evolutionary conserved regions (ECRs), and SINE-VNTR-Alu (SVA) at the FUS locus were studied as potential regulatory domains. In addition, the regulatory properties of the SVA at the PARK7 locus was also addressed.

1.8.10 Involvement of FUS/TLS in gene regulation

FUS/TLS is considered to be a multifunctional protein and has an association with multiple steps of gene regulation transcription and RNA splicing. From proteomic study it has been identified that the FUS/TLS is associated with specific and general factors in order to initiate the transcription (Law et al., 2006) also interact with nuclear hormone receptors (Powers et al., 1998). Added to that FUS can interact with gene specific transcription factors for example Spi-1 or NF-kappaB implicated in the spliceosome mechanism (Hallier et al., 1998, Uranishi et al., 2001). *In vitro* studies, meanwhile, demonstrate FUS/TLS is correlated with large transcriptional splicing complexes which bind to both the 5'and 3'of pre mRNA splice site. FUS/TLS is involved with several splicing factors, such as YB-1, serine-arginine proteins (SC35 and TASR) (Chansky et al., 2001, Göransson et al., 2002), hnRNP A1 and C1/C2 (Zinszner et al., 1994) FUS/TLS is involved in various mechanisms such as transcription and RNA processing for both mRNA and micro-RNA (Figure 1.7). For example FUS/TLS and TDP-43 have been found (by mass spectrometry) to play roles in micro-RNA processing by association with Drosha, the RNase III- type protein known for its function as a mediator of the first step of micro-RNA maturation (Kim et al., 2009) Figure 1.7. In addition, TDP-43 has been implicated in the cytoplasmic cleavage step of micro-RNA by Dicer complex, through its association with argonaute 2 and DDX17 proteins, which have known roles in these functions (Freibaum et al., 2010).

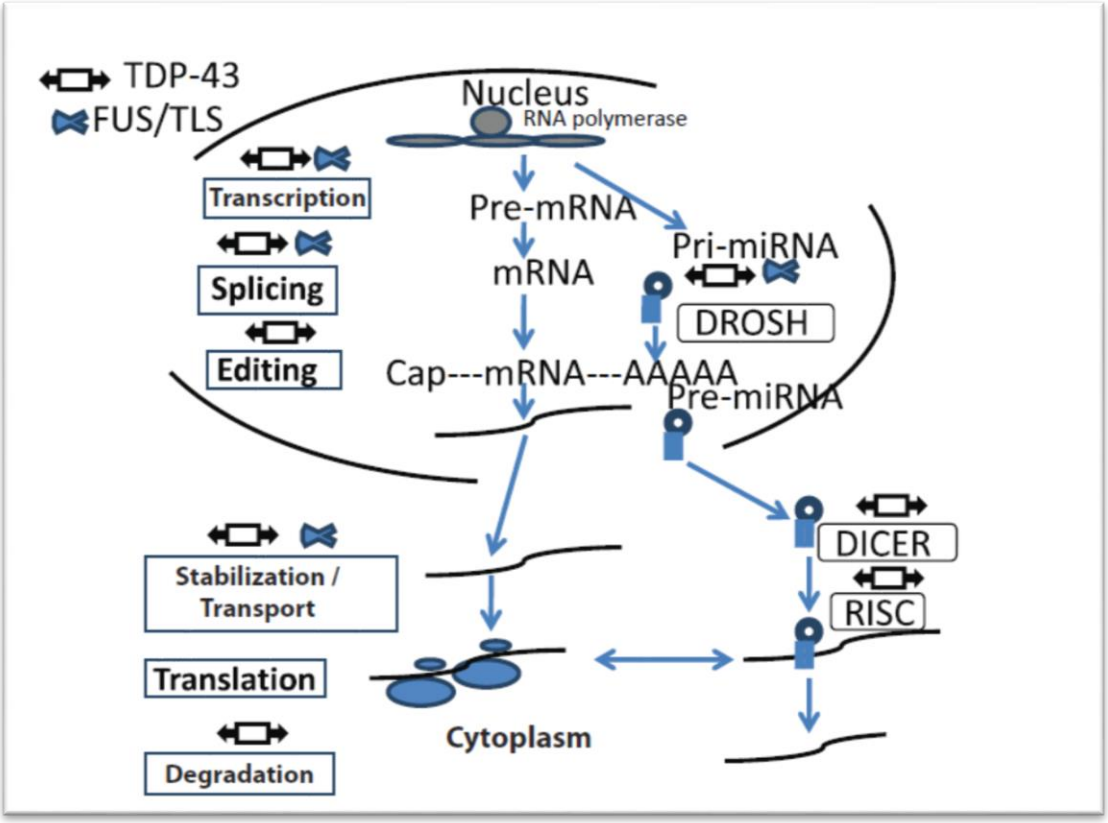


Figure 1.7: Schematic diagram demonstrating the involvement of the TDP43 and FUS/TLS genes in the metabolism of the RNA. FUS/TLS and TDP-43 both associated with DROSH and TDP-34 involved in Dicer. This Figure has been adapted from; Verma (2011)**.**

1.9 Elements that regulate gene expression.

Regulation of gene expression is a complex process and controlled spatially and temporally by interaction of signals that act on regulatory elements such as promoters and enhancers. These elements can recruit transcription factors which allow cells to control the rates of processes involved in gene expression such as chromatin decompaction, initiation of transcription and then elongation.

Previously, identification of the putative regulatory region in non-coding DNA was difficult for researchers because these regions were small and often widespread. The wealth of genomic data information now available from genomic browsers, however,

can accelerate identification using information about the genomic data of humans, rodents and pufferfish (fugu). Analysis of the potential regulatory element through genome homology serves to highlight sequences that might act as a putative regulatory element. Furthermore, these databases can hold information on the transcription factors across several cell lines (and species), along with the chromatin state, and can aid in the identification and analysis of the potential regulatory elements. The UCSC genomic browser (http://genome.ucsc.edu/index.html) is one useful and freely available application in the encyclopaedia of DNA elements, and can give a huge amount of information specifically about the functional elements of the human genome. This project is a collaboration between different research groups and is funded by the National Human Genome Research Institute (NHGRI). This tool is useful and can assist in identifying and analysing the putative regulatory elements in the genome.

1.9.1 Evolutionary conserved regions (ECRs)

Evolutionary conserved regions (ECRs) are regions of DNA sequences that have strong conservation across diverse vertebrate species. These ECRs can highlight important regions that might have a role in gene regulation (Drake et al., 2006) and can also assist in the discovery of novel genes (Pennacchio and Rubin, 2001). Exons are also strongly conserved between vertebrates, and some fraction of the non-coding sequences can have strong evolutionary conserved regions when observed between a ranges of species. These ECRSs that are located in the non-coding DNA sequence can pinpoint an important role of these sequences that has been conserved through evolution between a range of vertebrate species, for example human, mouse, fish (Lein et al., 2007, MacKenzie and Quinn, 2004). The potential role of the ECR can act as a

cis regulator for gene expression and can determine the spatial and temporal aspects of gene expression. In general, ECRs have been recognised as functioning as enhancers and repressors of gene expression (Pennacchio et al., 2006, Loots and Ovcharenko, 2007). Their role in gene expression has been identified and characterized *in vivo* by Pennacchio *et al.* 2006. As they tested 167 of a large group of non – coding elements from human chromosome 16 for enhancer activity in a transgenic mouse. These regions, are conserved between human and pufferfish or/and extremely conserved sequences in human –mouse-rat. Furthermore, 45% of the tested regions showed to act as enhancers at embryonic day 11.5. Furthermore, work by Paredes *et al*. 2011 studied the regulation role of the ECR from DRD4 gene, a key gene involved in regulating human social behaviour and cognition were demonstrated to act as transcriptional activity in primary cultures derived from rat cortex.

Identification of ECRs has been made much easier with aid of the ECR browser (Loots and Ovcharenko, 2007, Hillier et al., 2004). The ECR browser is a tool that has been developed to aid identification of novel genes, the distance of regulatory elements and putative transcription factor binding sites. In general, regulatory elements might be also located in different regions other than ECRs and might be resident in regions containing genetic variation. Genetic variation in human genome important as considered for differences between individuals of the same species. These polymorphic regulatory domains can have different form and categories for example variable number tandem repeats (VNTRs), transposable elements inserted into our genome.

1.9.2 VNTR (Variable Number Tandem Repeats)

VNTRs can be defined as a nucleotide sequence repeats greater than 6 bp that act as recognition motif for protein such as transcription factors and act as a regulatory domain. VNTRs always show variations in length between individuals (Haddley et al., 2008). VNTRs are considered to be a type of genetic variation and to have a role in gene regulation.

There are 600,000 VNTR candidates that are distributed in the human genome (Breen et al., 2008a). VNTRs can occur in functional positions, for example when they are located in splicing junctions between intron- exon and when they located within a first 50 base pairs of the promoters, enhancers and functional intronic regions. VNTRs can function when present in the recognition sequence to a protein DNA interaction, and provide various transcription binding sites through their repetitive features (Ferreira et al., 2008, Breen et al., 2008c, MacKenzie and Quinn, 1999).

VNTRs can have various copy number repeat, differing in their number of repeats between individuals. For instance, the serotonin transporter gene has a polymorphism within the intron 2, located in the promoter region. This polymorphism, consist of a VNTR, which differs in the number copies of repeats, containing 9, 10, and 12 copies of a 17 –bp elements (Heils et al., 1997). These variations in the number of repeat have been linked to unipolar and bipolar disorders. The 10 and 12 copies can act as strong positive transcriptional regulatory elements, in the hindbrain of E10.5 of mouse model(Heils et al., 1997). Meanwhile these two regions differs in the level of transcriptional activity *in vitro* as 12 copy polymorphism was showed to have stronger gene expression than the 10 copy polymorphism (Krueger, 2009).

The dopamine transporter gene (DAT1), meanwhile, has nine polymorphic versions of VNTRs located in the 3'UTR (Haddley et al., 2008). VNTRs have been shown to regulate gene expression, different alleles can drive expression in different locations and different alleles can respond to a stimulus differently *in vitro* and *in vivo* (MacKenzie and Quinn, 1999, Vasiliou et al., 2012, Roberts et al., 2007).

VNTRs have also been implicated in genetic predisposition to disease (De Luca et al., 2006, Munafò and Johnstone, 2008, Herman et al., 2005, Buhnik-Rosenblau et al., 2012). For example, the serotonin transporter gene (5HTT/SLC6A4) has two VNTRs, one located within the promoter region and the other located in intron 2. Thus the serotonin transporter gene has been implicated in alteration of the level of serotonin secretion between the synaptic cleft and neurones. Added to that the VNTR, which is located in the promoter region of the 5HTT gene has a long polymorphic region (LPR) and consists of two alleles, either 14 copies or 16 copies that have a 22=23 bp repeat, while the VNTR which is located within intron 2 has three different alleles that have 9, 10, or 12 copies and consist of 16-17 bp repeat (Stin2).

VNTRs have various regulatory classes but in general can share the important property that they are mostly located within non-coding sequences of the genome. VNTRs found in the genome emerge through evolution and are subject to evolutionary conservation between humans and primates, although they are not conserved in the lower mammals. They are implicated in the gene expression by possessing *cis*-regulatory elements or can have one or more transcription factor binding site. Many VNTRs are localized within the gene enriched region rather than non-genic regions, which supports the hypothesis that VNTRs can alter gene expression. VNTRs have been implicated in many diseases, such as diabetes, through the large non-coding

VNTRs located in the insulin gene (Ong et al., 1999) or depression (Caspi et al., 2003), attention deficit hyperactivity disease (Brookes et al., 2006), and addiction to cocaine (Guindalini et al., 2006).

1.9.3 Mobile DNA or transposable elements

Mobile DNA elements were first identified in 1950 by Barbara McClintock when studying gene regulation in maize (McClintock, 1950). Mobile DNA can be divided into two major classes; first major class the DNA transposons, which were, discovered in maize and are not actively mobile in humans and mice. They mobilise through a mechanism by which the DNA sequence that encodes for the transposon is cut out from its normal location in the genome and pasted into an alternative position. The second major class is the retrotransposons, this class are actively mobilised in humans and mice through an RNA intermediate by a mechanism called copy and paste, through which the retrotransposon is transcribed. In the transcription it will employ the intermediate RNA to act as a template for the synthesis of the cDNA by the RNA dependent DNA polymerase and this DNA can then integrate back into the genome and make a complete or a partial copy of the retrotransposon. The retrotransposons can be further classified into two subclasses: long terminal repeat (LTR) and non-LTR. This non-LTR class is still active in human genomes (Beck et al., 2011) and comprises long interspersed elements (LINEs), short interspersed elements (SINE) and SINE-VNTR-Alus (SVAs) and processed pseudogenes these transposable elements have recently become known for their activity in the human genome. Long interspersed elements (LINES) are the only elements that can move through the DNA. The majority of these elements are represented by the LINE-1 elements and comprise 18% of the human genome with about 500,000 copies (Goodier and Kazazian, 2008). LINE-1 has

two open reading frames (ORF) which encode for two proteins required for retrotransposition (Feng et al., 1996). This retrotransposition of the LINE1 can happen by a process called target primed reverse transcription (TPRT), in which this process itself can lead to 5'truncation, 3'transductions and target site duplications (TSD), internal rearrangement and inversions. LINE-1 have two important roles in the human genome; first they can cause expansion and proliferation and, second, they can provide mobility to non-autonomous retrotransposon elements such as SINEs, SVAs and processed or retro pseudogenes (RNA that has been produced by reverse transcription and inserted into the genome LINE-1 machinery by the process of 'copy and paste') (Pavlicek et al., 2006). Processed pseudogenes have no introns and no promoter and regulatory element in their structure but have a 3' poly A tail (Ding et al., 2006).

More than 120 families of Short Interspersed Elements (SINEs) have been found in the eukaryotic genome. They originated from the cellular RNA that has been transcribed by RNA polymerase III, which vary in length from 150-300 bp (Ding et al., 2006) The *Alu* element is the most popular form of SINE in the human genome. The *Alu* element is primate specific and has more than 1 million copies. These elements are named according to the *Alu1* restriction enzyme site in their sequences (Ullu et al., 1982). *Alus* are mobilized by the LINE-1 machinery and originate from the processed 7SL RNA gene. *Alus* and can regulate the transcription process by the internal RNA polymerase III (Batzer and Deininger, 2002).

SVA, specifically, is a non-autonomous retrotransposon that relies on the LINE-1 machinery to enter the genome. The SVA element is named from its main components of SINE, VNTR and *Alu* like sequences. Since SVAs are hominid specific and include a VNTR domain in their structure, it may be postulated that they have a role in genetic variation and involvement in the human genome. SVAs are the newest of the

retrotransposable elements and share the same characteristics as retrotransposon elements, for example they are flanked by target site duplications (TSDs) and terminate with a poly (A) tail. They can be inverted or truncated through their insertion in the genome since they can transduce 3'sequences through their transition in the genome. As a result, they are considered to be non-autonomous retrotransposons and they can transit through the genome by LINE-1 encoded proteins *in trans*. SVAs can be categorised into six SVA subfamilies (A,B,C,D,E,F) according to the SINE region, and their age. For example, the estimated age for the oldest sub type (A) is 13.56 million years (Myrs) while the youngest sub type (F) is 3.18Myrs old (Wang et al., 2005).

A canonical SVA consists of approximately 2 kilobases (Figure 1.8), however the size of the SVA insertion in the human genome may vary between 700 to 4000 base pairs. A canonical SVA consists of: starting from 5' sequences, a hexamer repeat 'ccctct' followed by sequences sharing homology to two antisense *Alu* like sequences, then a GC-rich variable number tandem repeat (VNTR), followed by a sequence sharing similarity with the *env* gene and a long terminal repeat of an endogenous retrovirus HERV-K10. It terminates with the canonical polyadenylation signal (polyA) tail. There are approximately 2700 copies of SVAs in the human genome and account for 0.13% of the human genome these may have been inserted in the human genome recently through SVA retrotansposons.

Figure 1.8: Schematic representing the structure of a canonical SVA. This image is taken from (Konkel and Batzer, 2010). The image shows the canonical SVA structure, consisting of target site duplications (TSDs), a full-length of elements containing from 5' to 3'a hexamer repeat (CCCTCT), a homology region of two antisense Alu sequences, a variable number of tandem repeat (VNTR) regions and a SINE region from HERV-K10 and human endogenous retrovirus, ending with a polyA-tail. Figure adapted from Konkel and Batzer (2010).

SVA elements have remained active in the human genome and play a role in cancer through various mechanisms by which SVA affects the regulation of gene expression. SVA, can mediate alternative splicing and might lead to the production of an abundant product, or can affect the level of mRNA production of a particular gene (Hancks and Kazazian, 2010)

1.9.4 The impact of the SVA on the human genome

Mainly, the insertion of the SVA has affected the gene regulation process as there are 2700 insertions of the SVAs in the human genome represent as a key feature for the polymorphisms and modulation of gene expression of specific genes that have a relationship with certain diseases. Moreover, SVA types have increased their number in the human genome at different stages of hominoid evolution, and their propagation has been associated with changes in human cognitive behaviour. (Vasieva et al., 2016) SVAs insertions have been linked with several disease, including, Duchenne muscular dystrophy, cystic fibrosis, haemophilia and several types of cancers (Hancks and Kazazian, 2010, Kaer and Speek, 2013). To date eight SVA insertions have been implicated with disease through several processes, such as exon skipping, insertions and decreases in the level of mRNA production.

1.9.5 Identification of the regulatory elements using the ECR Browser and UCSC Browser

The ECR Browser (http://ecrbrowser.dcode.org) is a dynamic tool that can provide the user with a wealth of information about the genomic structure of several compared species such as humans, dogs, rodents and fish, and can pinpoint important conserved regions. It is possible within the tool manually to manipulate the threshold and thus to generate, retrieve and display the data in a very easy way to define the sequence length of the base and compare the length and the percentage of conservation similarity. It can also provide the number and type of species whose genome has been compared with the desired sequence One of the useful application of this tool is that the generated data can be exported to the UCSC browser to identify the sequence homology between closely related species, such as humans and mice through using the BLAST tool or by using BLAST to search for distant species, for example rodents and fish (Paredes et al., 2011).

1.9.6 Development of a chick model for identification of the regulatory elements that regulate FUS and PARK 7 genes implicated in motor neurone disease

Many studies have advanced current knowledge by attempting to interpret the genome and characterise the varied associations of gene expression patterns with disease. In order to analyse the transcriptional activity of gene flanking sequences, it is essential to use cellular-based *in vitro* (Tanaka et al., 2010) (Trojsi et al., 2013)or whole organism-based *in vivo* models (Farley, 2013). Initially, *in vitro* models such as clonal cell lines or isolated primary cells were used to analyse the transcription activity of gene flanking regions. Typically, the potential regulatory element is cloned into a suitable plasmid vector along with a marker or reporter gene in order to visualise or quantify the effect of the element on transcriptional regulation. In cell line assays one of the most frequently used reporter genes is luciferase. Using this approach allows rapid analysis of regulatory elements to generate a robust data set and gives quantitative results on the relative efficiency of different sequences. *In vitro* models, however, do not reflect the tissue specific, spatial, or temporal gene regulation observed *in vivo*. Since many genes involved in human diseases show tissue-specific expression, *in vitro* cell lines might lack the relevant factors for tissue-specific gene expression, e.g., the human brain has a three-dimensional structure comprised of different cell types that would be difficult to recapitulate in a cell culture dish. Furthermore, neuronal interactions occur between different brain regions and include different cell types, and these interactions might have implications for gene expression. *In vitro* cultures containing only a few cell types in a monolayer, therefore, may not be an appropriate model (Uchikawa, 2008, Tanaka et al., 2010).

To circumvent some of the disadvantages of cell line models, *in vivo* animal models have been developed to address gene expression in an appropriate anatomical location for comparison with that seen in humans. For a suitable *in vivo* model certain criteria must be considered follows: (1) duration of life span, (2) wealth of knowledge about the organisms and degree of genomic comparison with the organisms that they are to be compared to, (3) ease of manipulation of the tissue, organs and DNA, (4) cost. A wide range of organisms have been started just from using a single cell, such as yeast and bacteria through to multicellular organisms such as *C. elegans, D. melanogaster*, fish, avian species, rodents and primates.

1.9.7 A mouse model can be used as an experimental model.

Mouse models can offer several advantages for the investigator. There is a well-established technique and a well-documented protocol for manipulating mouse embryos (Sambrook et al., 1989) and the mouse genome has been entirely sequenced (Chinwalla et al., 2002, Church et al., 2009). Mice have a brain structure and cell types that are similar to human (Bier and McGinnis, 2004). There are also a number of disadvantages, however, generating the transgenic mice protocol is: (1) Complex, (2) time consuming and cost 20 time more than chick model, (3) transgenic mice have a long life cycle and are slow to develop.

1.9.8 Advantages of the chicken embryo in vivo model compared with murine models.

The generation of a transgenic mouse protocol was found to be quite complicated, time consuming and labour intensive. In contrast, chicken embryos have many advantages that can be employed in order to identify and characterise the Cis regulatory element of the gene of interest (Kondoh et al., 2004, Uchikawa et al., 2003, Matsumata et al., 2005, Uchikawa, 2008). Fertilised eggs are cheap and can be obtained easily from the commercial hatcheries which produce fertilised eggs in large quantities for farming or pharmaceutical industry. Maintenance costs are very low as there are no animals to feed and animal houses to run equipment for. Chicken embryos have some disadvantage for use as an animal model instead of a mouse model, however. For example, the chick genome (International Chicken Genome Sequencing, 2004) has not yet been as well characterised as the mice genome as mice have been used more extensively by scientists. In the past, analysing gene regulation using chicken cells was only done through transfection of tissue cultured cells. This situation has now changed, however, with the successful application of the electroporation technique. *In ovo* electroporation can be defined as a technique to introduce DNA into living chicken embryos, thus potentially manipulating gene activities with high efficiency (Funahashi et al., 1999, Muramatsu et al., 1997). This technique has opened up a new field for the study of gene regulation (Uchikawa et al., 2003).

In order to predict the potential regulatory sequences, the application of the genomic comparison of the chicken and other species is used frequently (Uchikawa, 2008). One of the useful elements of the genome project is the ease of comparison of the entire genome sequences between vertebrate species, with the regulatory regions always able

to be recognised as non-coding DNA sequence blocks that show a very strong conservation among vertebrate species (Hardison, 2000). It is accepted that chickens are at an appropriate distance from mammals on the phylogenetic tree. This means that chicks and mammals can share strong conserved regions (Khokha and Loots, 2005), In addition, chickens and mammals belong to the same amniotes and might share various regulatory elements. The study of the gene regulation elements in chicks, therefore, can highlight important and universal paths that are conserved between vertebrates.

1.9.9 The electroporation technique can used in order to define the regulatory regions of a gene using reporter gene constructs.

Many methods have been used to transfer the reporter gene vector to cells or tissues in chicken embryos. Some of these methods are viral based, for example retroviral vectors have been used to identify genes that have a major role in the development and differentiation of myocytes in chicken embryos (Muramatsu et al., 1997). Viral methods are a typical approach for generating stable gene expression, while non-viral methods are considered only when transient gene expression is the aim. There are different transient transfection methods for chick embryos *in-ovo. In ovo* lipofection (LP) is a suitable method for gene transfection and shows expression in transfected tissues with a reporter gene vector. In addition to the LP method, there is another method called micro particle bombardment. This method has been used constantly in gene transfection in tissues of various animal species *in vivo.* Finally, electroporation or electro-transfection is a very simple and popular method, which can work on most types of cells, is cheap, and can direct the cloning to a particular location for example the forebrain and neural tube. Electroporation involves a sequence of electrical pulses

that allows cells to take up the DNA. This was successfully applied to chick embryos by (Neumann et al., 1980). With this method it is possible to observe reporter gene expression only 2 hours after transfection of DNA into chick embryos by *in Ovo* electroporation. A comparison study between the three mentioned non-viral protocols (MPB, LP and EP) was conducted by (Muramatsu et al., 1997) based on the expression intensity and the transfection efficiency of a Lac*Z* reporter gene in chick embryos *in ovo*. The EP technique was found to have the strongest reporter gene expression among the three methods and, indeed, EP scored a peak transfection rate of 50.0% (15/30) and a peak survival rate of 53.6% (30/56) (Muramatsu et al., 1997). Electroporation techniques have many advantages that make them the most suitable approach to be employed for *in ovo* gene transfection to chicken embryos. These advantages includes: (1) Simple and cheap, (2) transfection that can be done directly to the specific tissue, for example, neural tube, or brain, (3) it has become possible to transfect DNA to the cells in deeply located target tissue, for example motor neurone cells that are located in the ventral part of the neural tube (Timmer et al., 2001), (4) electroporation has been successful in achieving transfection of DNA into most cells and it can be done *in-ovo* in the developing chicken embryos (Muramatsu et al., 1997, Yasuda et al., 2000). In fact all three non-viral methods were first used in the 1990s and have been developed continually, but the development of electroporation techniques have superseded the others to a large degree (Uchikawa et al., 2003, Uchikawa et al., 2004, Burke et al., 2005, Tanaka et al., 2010, Farley, 2013). For these reasons, electroporation was considered to be the most appropriate transfection method for use in this thesis.

Electroporation entails the application of electric field pulses to cells or tissues. This is known to form some type of structural rearrangement to the cell membrane. These rearrangements consist of two main stages, the electroporation or electropermeabilisation stage, during which the cell's membrane produces temporary aqueous pathways, called pores, followed by the electrophoresis stage, which provides a dynamic force that causes the DNA, ions and molecules to be transported through these pores. Many theories have attempted to explain the events that occur within the membrane during the electroporation. For instance, the electric field should produce a potential difference of a critical value across the cell. After this critical value has been reached, it facilitates the molecules' entry across the cell membrane. Another theory is that the electroporation might raise the surface tension that leads to a change in the balance of two opposing forces of tension and line tension. Since surface tension can cause pore formation the line tension works as an opposing force and can overcome this pore formation (Weaver, 1995). Another theory claims that the electroporation creates an electric field that has the effect of compressing the membrane bilayer. While others believe that the polar head of the phospholipids is reorganised, which might cause awakening in the hydration layer (Chen et al., 2004, Ryttsen et al., 2000). Although the complete mechanism has not yet been elucidated, however, these theories do demonstrate that electroporation is asymmetrical, with the maximum effect being on the sides facing the electrodes and being most marked on the side facing the positive electrode (Hibino et al., 1993).

To produce electroporation there is a critical value, at an exact point on the cell. This critical value is normally between 200mV-600mV (Teissie and Rols, 1993, Gabriel and Teissie, 1997). Moreover, the electric field required to reach the critical value

works inversely according to the cell size; for example, small cells like bacteria require up to 200KV/cm whereas large cells only need 200V/cm (Teissie and Rols, 1993). Electrically connected cells such as those in a developing chick neural tube (Sheridan, 1968) perform electrically as if they were one big cell thus decreasing the electrical energy necessary for electroporation to 15-20 V. Beside the electric field strength there are another two elements, the duration and number of electric pulses, which regulate the electroporation level, elevating the field strength and increasing the area of the electroporation (Gabriel and Teissie, 1997). Too great a field strength will result in irreversible electroporation or joule heating, both resulting in cell death. A square wave pulse is now more favoured than that with an exponential decay as it reduces the amount of time thee electric field is applied and thus reduces the joule heating and risk of damage, i.e. burning (Uchikawa et al., 2003). Electroporation's disadvantages stem mainly from the nonspecific transport, of for example, calcium and sodium into the cell and, for example, potassium, out of the cell when the membrane has been permeabilised. In addition, there may be incomplete penetration such that only some cells are electroporated.

1.10 Aim of the project.

- The aim of the project was to establish a chick embryo based model in order to address the regulatory elements of gene expression in the form of ECR, VNTR, and SVA.
- In this thesis we test the hypothesis that DNA sequences that regulate FUS gene expression in motor neurones might be have a crucial role in the development of motor neurone disease. Mutations within 3'UTR of FUS gene were found in patients with ALS (1.2%) (Sabatelli et al., 2013b). This might

result in elevation in levels of FUS protein being crucial for cell phenotype and initiation or progression of ALS (Mackenzie et al., 2010), since FUS over expression in a mouse model led to progressive degeneration of motor neurones (Mitchell et al., 2013). Our laboratory has identified two regulatory elements, ECR and SVA, within this gene and showed them to have strong activity *in vitro*. Therefore regions of noncoding sequences from FUS gene, might have a crucial role in the regulation of gene expression.

- PARK7 gene was also addressed in this project due to implication of this gene in neurodegenerative disease and it was identified a regulatory domain in the form of SVA which also show to have strong activity *in vitro*

Chapter 2

Material and Methods

2 Materials and Methods

All chemicals were supplied by the Sigma-Aldrich unless indicated otherwise. Kits were supplied by manufacturers as indicated in the text.

2.1 Bacteriological Media

2.1.1 Luria Broth

25 g of Luria broth (Fluka Analytical) base was added per litre of distilled water and autoclaved at 121 ºC for 15 minutes.

2.1.2 Luria Agar

40 g of Luria broth (LB) (Fluka Analytical) Agar base was added per litre of distilled water and autoclaved at 121 ºC for 15 minutes, then cooled to 55 ºC. 10 mL of ampicillin stock was added. The LB agar was then poured into petri dishes (~25 mL per 100 mm plate).

2.1.3 Antibiotic

	Stock	**Working concentration**
Ampicillin	100µg/ml	50µg/ml

2.2 Electrophoresis Buffer for Agarose Gel Electrophoresis

5X Tris-Borate-EDTA.

Tris (TRIZMA base)	108 g
Boric acid	55 g
EDTA	5.84 g
Distilled water	Volume made up to 2 litres.

pH was adjusted to 8.3 if required.

T E buffer

10 mM Tris-HCl, pH 8.0, 1 mM EDTA.

Elution buffer

50 mM Tris-HCl, pH 8, 1 mM EDTA, 1% SDS, 50 mM $NaHCO_3$, 2.5 M NaCl.

2.3 Sources of cell lines

SK-N-AS Human neuroblastoma cell line, CRL- 2137 from the American Collection of Cell Culture.

2.4 Complete media for SK- N- AS cell line

Human neuroblastoma SK- N- AS cells were maintained in Dulbecco's Modified Eagles medium (Sigma, D5672), with 4500 mg/L of high glucose medium, 10% (v/v) foetal bovine serum (Thermo Scientific/Hyclone), 100 units per ml of penicillin, 100 µg/ml streptomycin (Sigma P0781), 1% (v/v) non- Essential Amino Acids (Sigma, M7145) and 200 mM L-glutamine (Sigma, D7513), in 5% CO_2 at 37 ºC

2.5 Freezing media

90% Foetal bovine serum, 10% DMSO.

2.6 Millonig's Phosphate Buffer

Sodium phosphate, monobasic $NaH_2PO_4.H_2O$ 12.68 g/L

Sodium phosphate, dibasic Salt NaH_2PO_4 43.8 g/L

$NaH_2PO_4.H_2O$ and NaH_2PO_4 were dissolved in 900 ml of distilled water and pH adjusted with HCl or NaOH to pH 7.4 at room temperature. The volume was made up to 1 litre with distilled water.

2.7 Millonig's Phosphate Buffer pH 7.4 (0.125M)

0.4 M Phosphate Buffer Diluted (Millonig's: water, 3:7)

2.8 4% Paraformaldehyde

Distilled water	35 ml
Paraformaldehyde	2 g
NaOH	4-5 drops
Phosphate Buffer 0.4 M	15 ml

35 ml of distilled H_2O was added to 2 g paraformaldehyde and stirred constantly whilst being heated gently to 65 ºC. 5 M NaOH was added dropwise until the solution cleared. When the solution had cooled to room temperature the volume was made up to 50 ml using 15 ml of 0.4 M Phosphate Buffer. pH was assayed and adjusted to pH 7. Ten ml aliquots were stored at -20 ºC.

2.9 Embryo staining

Sodium azide 10% stock solution

10 g of sodium azide in 100 ml of distilled water. This was stored at room temperature

2.10 Embryo staining solution

Phosphate Buffer	0.1 M pH 7.3
$MgCl_2$	2 mM
Sodium deoxycholate	1%

Potassium ferrocyanide	5 mM
Potassium ferricyanide	5 mM
NP-40	0.02%
X- gal	1 mg/ml (added prior to use).

2.11 Detergent Rinse

Phosphate buffer	0.1 M
$MgCl_2$	2 mM
Sodium Deoxy cholate	0.01%
NP-40	0.02%

2.12 Diluent solution

1% BSA,

0.12 M (Diluted Millonig's)

0.1% Triton

0.1% NaN_3

2.13 Analysis of endogenous gene expression

2.13.1 Tissue extraction

Chick embryonic day 5 (E5) neural tube, E15 forebrain and E17 cerebellum were used to purify total RNA. Chick eggs were incubated for 5-17 days at 38 °C. The egg shell was cut in order to obtain the embryo which was placed on a petri dish. The petri dish was then placed under a dissecting microscope and the neural tube tissue and the forebrain region of the developing embryo or the cerebellum was separated from the rest of the chick embryo. The dissected tissue was sliced into pieces less than 20 mg,

and then further sliced into slices less than 5 mm thick. This step was done as quickly as possible to prevent RNA degradation and was performed at room temperature.

2.13.2 Purification of total RNA from Animal Tissue

The RNeasy Mini Kit (Qiagen, Germany) was used to extract total RNA from the chick E5 neural tube, E15 forebrain and E17 cerebellum.

After harvesting, the tissue pieces were submerged completely in a collection vessel containing 600 µl RLT buffer. 10 µl of reducing agent β-mercaptoethanol (β-ME) was added to 1 ml of RLT Buffer prior to use. The RLT buffer consists of a highly denaturing guanidine isothiocyanate. This buffer can inactivate RNases in order to facilitate isolation of intact RNA. In order to disrupt and completely release all RNA contained in the sample, a mortar and pestle was used initially and subsequently the sample was homogenised using a handheld glass homogenizer. Following homogenisation, the lysate was passed through a 23-gauge needle fitted to an RNase-free syringe five times. The lysate was then centrifuged for 3 min at 8000 x g (10,000 rpm).The supernatant was transferred into a new microcentrifuge tube and one volume of 70% ethanol was added to the lysate and mixed immediately by pipetting.

700 µl of the sample was transferred to an RNeasy spin column placed in a 2 ml collection tube and centrifuged for 15 s at 8000 x g (10,000 rpm). The RNeasy spin column was washed with 700 µl of Buffer RW1 to remove the contaminants from the sample and centrifuged for 15 s at 8000 x g. The spin membrane was then washed twice with 500 µl of Buffer RPE and centrifuged for 15 s at 8000 x g to remove the contaminants; in the second wash, however the centrifugation was increased to 2 minutes. RNA was eluted by pipetting 30-40 µl of RNase–free water directly onto the RNeasy membrane, then centrifuging for 1 min at 8000 x g.

2.13.3 Design of primers for detection of mRNA expression

PCR was used to identify the expression of the FUS, ELP3, KIFAP3, and PARK7 genes, the primers were designed to bridge adjacent exons in order to avoid/distinguish any PCR product from contaminating gDNA which had acted as the template. The whole gDNA sequences of FUS, ELP3, and KIFAP3 PARK7 were taken from the UCSC genome browser (Hg19) (http://genome.ucsc.edu/index.html).

The exonic sequences were then used to design PCR primers using the Prime3 programme (http://frod.wi.mit.edu/). The potential set of primers were analysed in the in-silico PCR tool of the UCSC genomic browser to confirm they were unique to the target region of interest and that there were no other PCR amplicons. Primers were produced according to various criteria such as the product size, Tm, length of the primers, non-complementarity at 3' ends (Table 1 Appendix). Primers were synthesised by Eurofins (MWG primers, Eurofins Genomics Anzinger Str. 7a 85560 Ebersberg Germany).

2.13.4 First strand synthesis of cDNA from total RNA

cDNA was produced from the total RNA extracted in 2.2.1.4 by using either the Cloned AMV First-Strand cDNA Synthesis Kit (Invitrogen) or the Go Script Reverse Transcription Kit (Promega). The manufacturer's protocol was followed to convert up to 5 µg of total RNA to first-strand cDNA.

The following components were mixed and briefly centrifuged before use and then combined in a micro centrifuge tube.

Table 2.1: Component and the volume used in the reaction mix prepared for each sample.

Component	Volume
Primer(50µM Oligo(dT)$_{20}$	1 µl
10 mM dNTP mix	2 µl
DEPC-treated water	to 12 µl
RNA (5 µg)	(The volume of RNA varied between samples depending on the concentration of RNA)

The mixture was denatured by incubation at 70 ˚C on a heat block for 5 min and chilled immediately in ice water for 5 min, then centrifuged for 10 s and stored on ice. This allowed the oligo dT primer to anneal to the polyadenylated RNA and act as a primer for the subsequent synthesis reaction. The following reaction mix was prepared for each sample.

Table 2.2: The below reaction mix was prepared for each sample.

Component	1 reaction
5x cDNA Synthesis Buffer	4 µl
0.1 M DTT	1 µl
RNaseOUT ™ (40 U/ µl)	1 µl
DEPC-treated water	1 µl
Cloned AMV RT (15 units/ µl)	1 µl

Total Volume	8 µl

8 µl of the mix was pipetted into the reaction tube containing the annealed oligo 20, RNA and dNTP mix that had been prepared previously to make a final volume of 20 µl of reaction mixture, this was done on ice. The reaction tube was then transferred to a thermal cylinder preheated to 50 ˚C (specific temperature for AMV RT enzyme) and incubated for 60 min. The termination of this reaction was caused by heating at 85 ˚C for 5 min. The cDNA samples were then stored at -20 ˚C.

Alternatively, the Go Script Reverse Transcription Kit (Promega) was used to convert up to 5 µg of total RNA into first-strand cDNA. The following components were mixed and briefly centrifuged before being used and combined in a PCR centrifuge tube.

Table 2.3: The primer and RNA mix was prepared for each sample.

Component	Volume
Experimental RNA (up to 5 µg/reaction)	1 µl
Primer [$Oligo_{15}$ (0.5 µg/reaction)	1 µl
Nuclease-Free Water	3 µl
Final volume	5 µl

The primer and RNA were incubated at 70 ºC on a heat block for 5 min and then immediately placed on ice for 5 min. The mixture was then centrifuged for 10 s and stored on ice until the reverse transcription mix was added.

Reverse transcription reaction mix was prepared and combined on ice as in Table 2.4 below.

Table 2.4 Reverse transcription reaction mix was prepared and combined on ice.

Component	Volume
Go Script™ 10X Reaction Buffer	2 µl
$MgCl_2$ (25 mM)	4 µl
PCR nucleotide mix (10 mM of each dNTP)	2 µl
Go Script™ Reverse Transcriptase	0.6 µl
Recombinant RNasin®	0.5 µl
Nuclease free water	5.9 µl
Final Volume	15 µl

The 15 µl of reverse transcription mix was combined with 5 µl of RNA and primer mix and then incubated on a heat block first for 25 ºC for 5 min, and then at 42 ºC for up to one hour. The reaction could be stored at -20 ºC or used for PCR immediately.

2.13.5 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was performed to amplify cDNA for analysis of gene expression using Go Taq Flexi DNA Polymerase (Promega) and utilising the cDNA that was generated in (2.2.1.4) as a template (Table A1 Appendix)

Table 2.5 The components and reaction volumes used in the PCR reaction Taq polymerase was used in PCR reaction

Component	**Volume**
10X PCR Buffer minus Mg	5 µl
$MgCl_2$(50 mM)	1.5 µl
PCR nucleotide mix (10 mM of each dNTP)	1 µl
Upstream Primer (10 µM)	1 µl
Downstream Primer (10 µM)	1 µl
Taq polymerase (5 U/µl)	0.4 µl
Nuclease free water	38.1 µl
cDNA template	2 µl
Final Volume	50 µl

Table 2.6: The following PCR program was run for amplification of target cDNA and optimising the annealing temperature for each primer set

	Activity	Time	Temperature
Step 1 (1 cycle)	Denature	2 min	94 ˚C
Step 2_(30 cycles)	Denature	1 min	94 ˚C
	Anneal	1 min	55 ˚C
	Extend	1 min	72 ˚C
Step 3	Final incubation	10 min	72 ˚C
Step 4	Hold	Indefinite	4 ˚C

Gradient PCR was performed and optimising annealing temperatures were ranged between (55 ˚C - 65 ˚C) with a specific band obtained at 62 ˚C, which was performed for the PARK7 primer set specifically. PCR was run in a peqSTAR 2X (peqlab) Gradient Thermocycler.

2.13.6 Polymerase chain reaction using a proof reading polymerase

PCR was performed for amplification of the DNA template using Phusion High-Fidelity polymerase enzyme. Restriction enzyme sites were incorporated at the 5’ end of each primer in order to facilitate directional cloning. The Phusion DNA polymerase master mix was prepared as follows:

Table 2.7: The components and reaction volumes used in the PCR reaction Phusion polymerase was used in PCR reaction

Component	Volume
5 X Phusion High – Fidelity Buffer	10 µl
10mM dNTPs	1 µl
10µM Forward primer	2.5 µl
10µM Reverse primer	2.5 µl
Template	X µl
Phusion polymerase	0.5 µl
dH_2O	X µl

Table 2.8: The following PCR program was run for amplification of target DNA

	Activity	Time	Temperature
Step 1(1 cycle)	Initial denature	1 min	98 ºC
Step 2 (25 cycle)	Denature	5 s	98 ºC
	Anneal/extension	20 s	72 ºC
(1cycle)	Final extension	10 min	72 ºC

2.14 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments by size or for restriction digests analyses. By this process, negatively charged nucleic acid molecules move through an agarose matrix within an electric field. The shorter the molecule the faster it can move through the matrix, thus migrating further than longer molecules. Since DNA is a negatively charged molecule it moves towards the positive terminal of the electric field.

Agarose gel 1-2% was made by dissolving an appropriate amount of agarose (Bioline) in 0.5X TBE in a conical flask, the flask was heated in a microwave oven and then removed and the solution mixed before being returned to the microwave to boil. Once the agarose was completely in solution it was cooled at room temperature for 5 min and then the nucleic acid stain was added to allow UV visualisation (either 1 µl of GelRed per 10 ml (Cambridge Bioscience) or 0.5 µl per 10 ml of ethidium bromide (Sigma 10 mg/ml) was added). The percentage of agarose gel used was dependant on the size of the DNA fragments that were to be resolved on the gel. The gels were poured into casting trays and combs were placed inside. The gel was allowed to set at room temperature for 40 min, and, once set, the comb was removed. The gels were placed in a gel tank filled with 0.5X TBE running buffer which was part of the electrophoresis apparatus. The addition of 5 µl of ethidium bromide was recommended for every litre of running buffer if ethidium bromide was used as a nucleic acid stain. Prior to loading the sample into the wells, 1 µl of loading dye (Promega 6X) was added to each 5 µl sample, this was to enable monitoring of how far the DNA sample migrated. Depending on the expected DNA fragment size on the gel a DNA ladder, either 100 bp (Promega) or 1 kb (Promega) was loaded, usually in the first gel well.

After the samples were loaded into the wells the lid on the gel apparatus was attached ensuring the positive (red) terminal was at the bottom of the gel as the voltage was applied. The voltage was set at a standard 5 V/cm, with the length of exposure to the current being dependent on the DNA fragment size and the percentage of the gel. The DNA was left for at least 1 hour to migrate through the gel towards the positive terminal. After the electrophoresis was done any bands were visualized on a UV transilluminator (BioDoc-it Imaging System) and photographs were taken using a digital camera.

2.15 Methods for cloning

2.15.1 Designing primers for detection of endogenous FUS, KIFAP3, ELP3, PARK7 genes expression from purified total RNA.

Primers that could be used to amplify the required sequences of chick genome DNA of the gene of interest were designed using Primer3 (http://frodo wi.mit.edu/). The primers were designed to bridge exon boundaries in order to distinguish the product from any other amplicon arising from chick gDNA contamination. First, the chick gDNA of the gene of interest was obtained from the UCSC genomic browser (H19), then the intron sequences were deleted. The obtained sequences were applied to Primer3 software which is available from either the UCSC Genome Browser (http://genome.ucsc.edu/index.html) or from the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The generated primers were based on several criteria such as the melting temperatures, usually between of 50 °C and 65 °C and a, GC% content between 40-60%, and then these primers were applied to the *in silico* PCR tool of the UCSC genome browser for the determination of the primer specification for the area of interest only, and no other potential PCR product

2.15.2 Bioinformatic Analysis

Evolutionary conserved region (ECRs) potentially containing regulatory elements were identified by genomic comparison using the ECR Browser (http: //ecrbrowser.dcode.org (Ovcharenko et al. 2004)) and the UCSC Genomic Browser Hg 19 (http://genome.ucsc.edu/index.html).

The ECR Browser is a dynamic tool demonstrating multiple alignments for genomic sequences of several vertebrate species including humans, rodents and fish.

The ECR Browser was set to highlight only sequences with a length of more than 100 bp with at least 70% similarity across the species sequences. In addition, the data was exported to UCSC Browser and used to search for sequence homology between closely related species, for example, humans and mice, using the BLAST tool, or to compare sequence homology between more distant species such as rodents and fish (Paredes et al., 2011). The human FUS gene was compared with chimpanzees, dogs, rats and mice using 70% identity and a length of 100 bp as the parameter to identify conserved sequences. Of the conserved regions, a region 4 kb upstream and downstream was analysed and coordinated and were obtained from Hg19 build. Typically, regions of interest were identified and were extended by 100-200 bp on both sides to facilitate primer design. Primers were designed incorporating restriction enzyme sites at the 5' ends to facilitate construction of the plasmids.

2.15.3 Restriction enzyme digests

Restriction enzyme digests were used to make the specific nucleic acid overhangs which allowed the PCR product to be ligated efficiently into the vector in the correct orientation, or they were used as an indicator for the presence of the insert and to confirm the direction of cloning. Restriction enzymes from both Promega and from New England Biolabs were used (approximately 5 unit/1 µg DNA).

The components of reaction used with restriction enzymes from Promega or New England Biolabs were as follows:

Nuclease free water	final volume of 20 µl
10X specific restriction enzyme Buffer	2 µl
Acetylated BSA (10 µg/µl)	0.2 µl
DNA (1µg)	Y µl
Restriction Enzyme (typically, 10 unit/ µl)	0.5 µl

The digestion reaction was incubated in a water bath for 1-4 hours at the recommended temperature and buffer that was specified for each enzyme's optimum activity. The reaction was then purified using a Promega mini column system cat. # A9282, and wash buffers, and was recovered from the column using an elution buffer. A fraction of the eluate was then run on a 0.8% agarose gel in order to visualise the DNA band size, by loading a DNA ladder (100 bp ladder; Promega, Cat. No. G2101 or 1 Kb ladder; Promega, Cat. No. G5711). This allowed confirmation that the reaction had worked or allowed gel purification of specific bands.

2.15.4 Extraction of DNA fragments from agarose gels

Agarose gel containing the required fragment of DNA was illuminated under long wave UV translumination. The DNA fragment was excised from the agarose gel using a clean blade. The DNA fragments were extracted using Promega Wizard SV Gel & PCR clean-up system.

2.15.5 SV Gel clean-up kit protocol

Promega wizard SV Gel & PCR clean-up system cat. #A9281 was used to extract the fragment DNA from the agarose gel. This kit has a special membrane binding solution (MBS) which can dissolve the agarose gel and aid the binding of the DNA to the membrane in the minicolumn assembly. A membrane wash solution is provided to eliminate contaminations and nuclease free water is used to elute the DNA from the membrane. Briefly, the excised gel slice was dissolved at a ratio of 1 µl MBS to 1 mg of gel at 50 ºC – 65 ºC or, for a PCR amplification, an equal volume of PCR reaction and MBS. The mixture then was centrifuged through a minicolumn assembly at 16,000 x g for 1 min to allow the DNA to bind to the membrane. The membrane was washed with MBS, centrifuging at 16,000 x g twice. Following the evaporation of the residual ethanol, the DNA was eluted from the membrane in 50 µl of nuclease-free water.

2.15.6 Ligation of DNA fragments into plasmid vector

Two cloning methods were followed, one allowed for directional cloning and the other non-directional cloning.

For directional cloning the inserts were cloned into the vector using the unique overhangs produced from two different restriction enzyme digests, located at specific ends of the insert. When non-directional cloning was used, the vector was treated with

Shrimp Alkaline Phosphatase SAP (Fermentas Catalogue Number EF0511) catalyses the release of 5'-phosphate groups from DNA, using the following protocol to reduce the risk of vector re-annealing.

2.15.7 SAP treatment Protocol for dephosphorylation of DNA 5'- termini.

17 µl of the solution containing the DNA sample, 10 X reaction buffer 2 µl, and SAP (1 µl at 1 unit/1 µl) were incubated at 37 ºC for 30 min before the reaction was stopped by heating at 65 ºC for 15 minutes.

2.15.8 Ligation reactions

Ligation reactions usually followed a molar ratio of vector to insert of 1:3. The following equation was usually followed in order to estimate the volume of insert required for ligation.

$$\frac{\text{ng vector x insert size kb}}{\text{vector size kb}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}} = \text{ng insert}$$

A typical ligation reaction contained:

Vector DNA	1 µl (100ng)
Insert DNA	Y µl
10X Ligation Buffer	1 µl
Nuclease free water	Z µl
T4DNA Ligase (1-3 U/µl)	1 µl
Final volume	10 µl

The ligation reaction was incubated for 3 hours at room temperature and then used for transformation of the chemically competent cells or stored at -20 ºC. Usually, two

control reactions were performed along with the ligation reaction; the first just with vector control, and the second with vector control in tandem with ligase control.

2.15.9 Transformation of chemically competent DH5α E.coli cells

A vial of chemically competent DH5α *E.coli* (Invitrogen) stored at -80 °C was defrosted on wet ice and 50 µl aliquots were placed in 1.5 ml microcentrifuge tubes. 10 µl of ligation mixture or 10 ng of plasmid DNA were added to the 50 µl of competent DH5α cells, gently mixed and then placed on ice for 30 min. Then the cells were placed in the water bath at 37 ºC for 20 s for heat-shock and then incubated on ice for 2 min. 950 µl of LB broth which had been pre warmed at 37 ºC was added to the transformed cells and they were then incubated in a shaking incubator at 37 ºC and 225 rpm for 1 hour. The tube containing the cells was positioned at an angle to assist gaseous exchange to ensure the mixture was aerobic. 50- 200 µl of transformed cells were spread on a pre warmed LB agar plate which contained the appropriate antibiotic (100 µg/ml ampicillin) and incubated inverted at 37 ºC overnight to allow growth of colonies of bacteria containing plasmid. The rest of the transformed mixture was stored at 4 ºC for future use.

2.15.10 Plasmid DNA isolation from bacterial cultures

To produce plasmids suitable for electroporation with a $_{260}$ nm: $_{280}$ nm ratio >1.7, overnight cultures of bacteria containing the plasmids were purified using a plasmid prep kit (Qiagen). The mini, midi and maxi version of their products were used depending on the amount of culture being purified. The manufacturer's guidelines were followed. Overall, the methodology comprises lysis of the bacteria to release the plasmid, recovery of the plasmid on a membrane, followed by elution to release the plasmid. In order to store the transformed bacteria for long periods of time, glycerol

stocks were prepared. 1.4 ml of a fresh overnight culture was transferred to a microcentrifuge tube and the bacteria were pelleted by centrifugation at 8000 rpm for 3 min at room temperature. The supernatant was discarded and the pellet was gently resuspended in 0.5 ml of sterile 15% glycerol (v/v in LB broth) and then transferred to a cryovial. This was immediately frozen and stored at -80 °C.

2.15.11 Mini-preparation of plasmid DNA

To purify up to 20 µg plasmid DNA for screening of colonies to identify the correct clones, the Miniprep protocol was followed in accordance with the manufacturer's guidelines (Qiagen, Cat No. 27106). Colonies were selected from the agar plates that the transformation mixture was cultured section (2.1.2*)* and the bacteria grown overnight in 5 ml of LB broth media containing the appropriate amount of antibiotic. This step will result in expansion of the bacteria containing the plasmid of interest. The 5 ml of the overnight culture was pelleted and then suspended in a 250 µl lysis buffer in order to lysis the bacteria to release the plasmid. Once the cellular lysis was completed and the plasmid was released, the reaction was neutralised and cell debris and RNA, proteins and low molecular weight impurities were removed with salt wash. Plasmid DNA was eluted in 50 µl nuclease free water or elution buffer and then collected into a fresh 1.5 ml microcentrifuge

2.15.12 Maxi-preparation of plasmid DNA

The maxiprep kit protocol was followed in accordance with manufacturer's guidelines (Qiagen, Cat No. 12165). It was designed to purify up to 500 µg of high plasmid DNA from E.coli cells cultured in 100 ml selective Luria broth overnight with supplemented appropriate antibiotic at 37ºC with shaking (225 rpm). The resulting DNA pellet was then resuspended in 200-500µl of a lysis buffer. Once the cellular lysis had been achieved and the plasmid was released the reaction was neutralised immediately and the cellular debris was removed. The plasmid was purified, washed and eluted through the column by gravity flow. The DNA was precipitated from the eluate by the addition of 0.7 volumes of isopropanol at room temperature. The DNA pellet was washed with 70% ethanol in order to remove the excess salt. The ethanol was removed and the DNA pellet was air dried and resuspended in 500 µl of elution buffer or nuclease free water.

The concentration of the plasmid DNA, purified by mini or maxiprep methods, was measured using either a spectrophotometer or a Nanodrop 8000 and then stored at -20 °C for working stocks or stored at -80ºC for long term storage. DNA purity ratio = $OD_{260}/OD_{280.}$ The recommended ratio for successful transfection was ≥1.7

2.15.13 Analyse of RNA, DNA Measurements, and protein contamination by using spectrophotometer

2.15.13.1 RNA measurement

The purity of RNA isolated by the RNeasy Mini Kit was measured by the ratio of absorbance readings at 260 nm with 280 nm in a spectrometer. DNA and RNA both absorb ultraviolet light with a peak in absorbance at 260 nm. Nucleic acids are commonly contaminated with different molecules, such as proteins, which have a peak absorbance of 280 nm. The spectrophotometer exposed the samples to ultraviolet light at 260 nm and a photo-detector measured the amount of light that passed through the sample. The spectrometer was then set to expose the sample to a wave length of 280 nm, the measurements of light that passed through the samples was recorded. The ratio of absorbencies was calculated by dividing the $_{260}$ nm reading by the $_{280}$ nm reading. Pure RNA has a 260/280 ratio of 1.9-2.1, meaning that the values expected for high quality RNA are ~2.0.

The concentration of the RNA was also calculated as shown below.

Concentration µg/µl = Absorbance$_{260}$ X **dilution factor** X **40 (co-efficient of RNA)**

An absorbance reading of 1 at 260 nm = 40 µg RNA/ml

The concentration of the RNA isolated determined the volume of each sample needed in cDNA synthesis. The samples of RNA were stored at -20 °C.

2.15.13.2 DNA measurement

DNA concentration was measured using the standard protocol of absorbance at 260 nm using a spectrophotometer. DNA concentration measurements were performed by diluting the DNA sample with distilled water to give a concentration in a range between 10 µg/µl and 50 µg/µl. First the spectrophotometer was calibrated at $_{260}$ nm using distilled water. Then, the optical density of the DNA sample concentration in ng/µl was obtained by OD at 260 nm and the final concentration calculated using following formula.

OD_{260} X 50 X dilution factor = concentration of DNA (µg/µl)

OD_{260} = optical density of sample at 260 nm.

An absorbance of 1 at 260 nm = 50 µg DNA/ml

The DNA purity was obtained at 260 nm (OD_{260}) and also at 280 nm (OD_{280}); the purity of the DNA was calculated using the following formula.

DNA purity ratio = OD_{260}/OD_{280}

OD_{260} = optical density of the sample at 260 nm.

OD_{280}= optical density of the sample at 280 nm

The recommended ratio for successful transfection was ≥1.7

2.15.14 Sequencing

Plasmid samples with the insert cloned in were sequenced externally using either Source Bioscience Sanger Sequencing service or Dundee DNA Sequencing and Services (see Figure 6.6 in the Appendix).

2.15.15 Reporter gene construction for in vivo analysis in a chick embryo model

2.15.15.1 Construction of FUS ECRs phrGFP Reporter gene constructs

The human FUS gene promoter fragments ECR PP (582) and PP (243) that were identified in (Chapter 3 section 3.3.12) were subject to PCR and cloned in pGL3b plasmid for use in the *in vitro* study. For *in vivo* analysis of the FUS promoter in chick embryos the addressed regions above were cut from their original pGL3b plasmid using *SacI* and *Bgl II* restriction enzymes. These reporter gene constructs were prepared by and a kind gift from Dr Thomas P Wilm, (University of Liverpool). Subsequently, these fragments were cloned into the *SacI/BamHI* site of the multicloning site of promoter-less vector phrGFP (Stratagene, UK) upstream of the GFP reporter gene. These clones were named FUS PP phrGFP and FUS ECR PP phrGFP (see Figure 6.11 and Figure 6.12 in the Appendix both these two plasmid constructs were constructed and prepared by Dr Vivien J Bubb and Christine Cashman at the University of Liverpool).

2.15.15.2 Construction of INT PP phr GFP and INT ECR PP phrGFP

In order to clone the putative regulatory element from FUS intron 1 into a location upstream of the FUS promoter sequence in ECR PP GFP and PP GFP, a 704 bp fragment was amplified by PCR using proofreading Phusion DNA polymerase (NEB, UK) and primers 5'-GATGAGATCTATGGCCTCAAACGGTAGGTAAGG and 5'-AGGTGCTAGCGAAAGAAATTTAGGCGGGAAAAACTCTCGGGC,

introducing *BglII* and *NheI* restriction sites for directional cloning. The resulting constructs were named FUS INT PP phrGFP and FUS INT ECR PP phrGFP, respectively (see Figure 6.13 and Figure 6.14 in the Appendix).

2.15.15.3 Construction of FUS SVA ECR PP phrGFP, FUS SVA PP phrGFP and FUS VNTR ECR PP phrGFP, FUS VNTR PP phrGFP

FUS SVA and FUS VNTR sequences were excised from the original pGL3P vector and then amplified by PCR using standard Phusion polymerase conditions. The primers used are outlined below and incorporate *NsiI* and *XbaI* restriction enzyme sites to facilitate directional cloning:

SVA UP 5'- TTGCATGCATGTGACTATTGCATACCTTGC-3' and SVA DN 5'-GACGTCTAGAGGAGAGGTTGTCATGGTACA-3', and VNTR UP 5'-TTGCATGCATCATCAGTTTTCCCTCAGACCCAG-3' and VNTR DN5'-GACGTCTAGAGTTGGGGGTAAGGTCACAGA-3'.

The resulting products were cloned into the *NsiI*/*XbaI* site of FUS PP phrGFP and FUS ECR PP phrGFP respectively. The result product created FUS SVA PP phrGFP, FUS

SVA ECR PP phr GFP and FUS VNTR PP phrGFP, FUS VNTR ECR PP phr GFP (see Figures 6.15, 6.16, 6.17, 6.18 in the section for chapter 4).

2.15.15.4 Generation of PARK7 SVA construct in forward orientation construct.

In order to clone the PARK7 SVA forward sequence into the reporter plasmid FUS PP GFP, the insert was released from the Zero Blunt PCR vector from Invitrogen with *Nsi1* a kind gift from Dr Abigail Savage at the University of Liverpool).

The reporter plasmid phr FUS PP GFP was digested using the same restriction enzyme *Nsi1* and then the linearized plasmid was cleaned over a Promega # A9282 mini elute column to remove all the buffers and then dephosphorylated with shrimp alkaline phosphatase. The released fragment was cloned into the *Nsil* site of the reporter plasmid FUS PP GFP and ligated to produce a construct named phr GFP FUS PP PARK7SVA (Figure 6.19 in the Appendix) and screened the minipreps with *BglI* as this enzyme cut once with the insert and twice with the backbone and released fragments 1229, 1724, 2697.

2.16 Injection and transfection by electroporation.

The most appropriate Hamburger-Hamilton (HH) stage of embryos to inject the neural tube was found to be at about HH 12-14. Earlier than this, the neural tube has not formed a closed central canal of sufficient length, whereas later than this the central canal of the neural tube has become too small. The protocols involving Chick embryos described in this thesis were from the Cold Spring Harbor Manual (Sambrook et al., 1989) .

2.17 Plasmid mixes for the injection

The plasmid mix contained a mixture of the two plasmids at a maximum concentration load of 6 µg/µl. The constitutively active td Tomato plasmid (a kind gift from Marco Marcello, University of Liverpool) at final concentration of 1 µg/µl and the phrGFP (human recombinant green fluorescent protein) reporter plasmid containing the sequence under test (at a DNA concentration adjusted between of 2-5 µg as per instruction adopted from Cold Spring Harbor Manual (Sambrook et al., 1989) were suspended in PBS +1 mM $MgCl_2$, containing 0.2% (v/v) of fast green (FG) to aid visualization during the injection process.

The injection of the td Tomato plasmid (Figure 3 in the Appendix) acted as a positive control for the transfection process and indicated the location of the reporter plasmid.

2.18 Accessing the Embryos

It was determined that eggs were normally at 12-14 HH, and thus at the correct stage for use, at 48 hr after incubation at 38 °C. The eggs were placed in a petri dish which had been prepared previously, with a blue tag button in the centre of the petri dish to hold the egg properly for windowing; the egg was then cleaned with tissue immersed with ethanol. Then the pointy end of the egg was punched with an egg punch and a syringe fitted with a 19G needle was inserted through the hole, directed below the equatorial mark to avoid damaging the yolk and used to withdraw 2-3 ml of albumen so as to allow the embryo to sink away from the area where the window was to be cut to avoid damaging the embryo in the subsequent process. The precise volume removed depended on the size and freshness of the egg, the fresher and bigger the eggs the greater the amount of albumen that was needed to be withdrawn. Then the puncher hole was taped up with a piece of laboratory tape (Shamrock brand) and the egg was

prepared for windowing. An oval window was drawn with a pencil on the top of the egg, centred on where the embryo should be (normally the embryo floated on the highest part of the egg during incubation). A suitable length of Sellotape Magic tape (2.5 cm wide) was placed over the marked oval. Using dissecting scissors, the marked oval was cut through the shell (Figure 1). The dust from the cutting was removed from the egg. A suitable length of the Sellotape Magic tape (2.5 cm wide) was placed over the partially cut oval or 'window'. The tape was stuck down on both its short and long ends leaving a piece raised up at one end to aid unsealing the window later. An observation of the egg was made to ensure that the embryo had developed to the appropriate stage (Figure 2.2) and that an adequate volume of albumen had been withdrawn; any bubbles that had possibly formed in the albumen were ruptured with a hypodermic needle. The window was then sealed and the Sellotape was pressed down with care all around to form a seal.

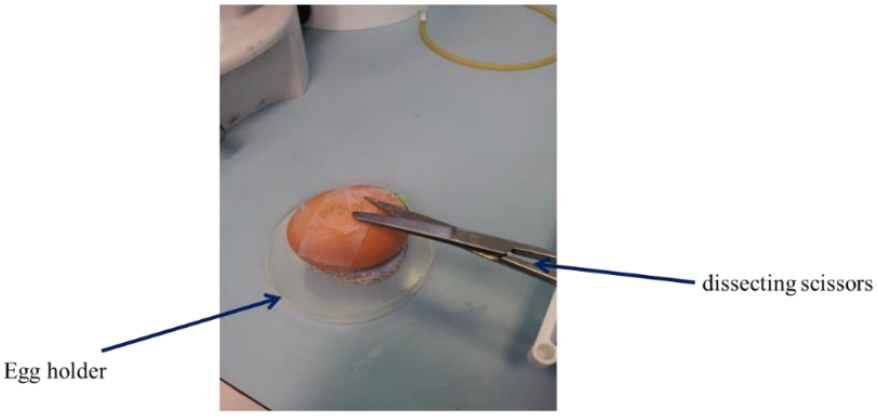


Figure 2.1: Egg being windowed with dissecting scissor. The chick egg was placed on the Petri dish and prepared for windowing with dissecting scissors.

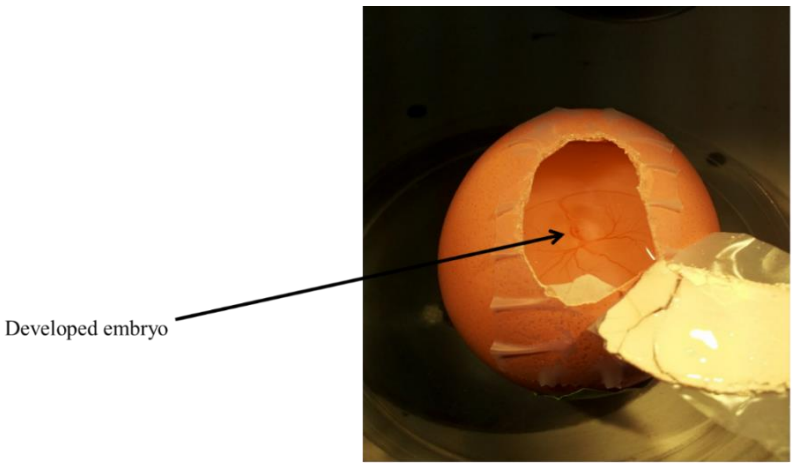


Figure 2.2: The windowed egg with the developing embryo visible. The windowed was placed under a microscope and prepared for micro injection.

2.19 pG-td Tomato

pG-td Tomato sequence was PCR-amplified from pG-td Tomato using the forward primer 5'ATAGGAATTCCGTGTACGGTGGGAGGTCTA-3' and reverse 5'-GGCCGTCGACATCATTTTACGTTTCTCGTTC-3'. These introduced the *Eco* RI and *SalI* restriction sites as indicated, for directional cloning into the plasmid pIRES eGFP. The pIRES eGFPcassette was removed from the vector using the *Eco* RI and *XhoI* restriction sites and replaced by the tomato reporter gene, such that the Tomato reporter was located downstream of the chick beta actin promoter (this plasmid was a kind gift from Dr Vivien J Bubb and Christine Cashman, University of Liverpool), see Figure 6.9 map of td Tomato. The td Tomato plasmid was co-injected with a promoter-less reporter vector phrGFP (Stratagene, UK).

2.20 IRES GFP

IRES containing GFP-expressing reporter plasmid was a gift from Dr Jon Gilthorpe (Department of Developmental Neurobiology, King's College London), see Figure 6.8 map of iresGFP. This is a CMV/ β-actin promoter upstream of the internal ribosome entry site (IRES) under the influence of the cytomegalovirus enhancer resulting in constitutive expression by the plasmid and high levels of expression of the, downstream, enhanced green fluorescent.

2.21 Marker gene

In order to demonstrate that transfection had occurred and was located at the site of the injection a marker gene was co-injected with the reporter gene. The marker gene plasmid can act as a positive control for a successful electroporation. Two marker genes were used in developing this thesis, the first marker gene used was IRES GFP,

and this was co-injected with a p1229-based vector, see Figure 6.7 Map of P1229, which is a modified version of pBLUESCRIPT, with the entire LacZ ORF/cDNA under the control of a human βglobin minimal promoter inserted into the polylinker area. This vector required an enhancer sequence that is cloned upstream of the promoter in MCS, and when the transcription factors bind the cloned enhancer sequence, the human βglobin minimal promoter is activated and the LacZ gene is transcribed. This results in the expression of the β-Gal by the cell. The β-Gal is later visualised by staining. These two marker proteins have been shown to be expensive in terms of both time and labour and hence were replaced with other marker proteins. IRES GFP was replaced with pG-td Tomato (a kind gift from Dr Marco Marcello, University of Liverpool) plasmid and this was used as a marker gene for the remaining experiments and the p1229 vector was replaced with the no promoter phrGFP reporter gene system.

2.22 Injection of the plasmid

The egg was placed under a dissecting microscope and the window shell removed (Figure 2). The vitelline membrane was removed using two forceps. A micropipette of borosilicate capillary glass tubing, with a filament, (Warner Instruments, G150TF-3) was pulled (Sutter Instruments Model P-87 Flaming/Brown pipette puller) and attached to a plastic mouth pipette aspirator tube assembly for calibrated microcapillary pipettes (Sigma). The sealed end of the pipette was snapped off to give a needle of the appropriate diameter. These were then loaded with 100 – 200 µl of plasmid mixture (Figure 3). The neural tube or brain of the chick embryo, as appropriate (Table A2 Appendix), was penetrated with the micropipette tip (Figure

2.3) and the plasmid mix was inserted into the central canal of the neural tube or the lumen of the brain via a small volume of buccal pressure.

After the micropipette was removed, gold plated electroporator straight wire electrodes of 3 mm or 5 mm length (Harvard Apparatus, Inc. BTX Models 514 (3 mm) 512 (5 mm)) (Figure 2.4) were used for electroporation of the chick embryo neural tube (Figure 2.5) while paddle electrodes (Figure 2.4) were used for electroporation of the brain region, specifically the mid brain of the chick embryo (Figure 2.6) were positioned on either side of the embryo with a gap of 5 mm and attached to the electroporator. A drop or two of Phosphate-Buffered Saline (PBS) (Thermo Fisher Scientific) was placed on the electrodes to increase electrical connection.

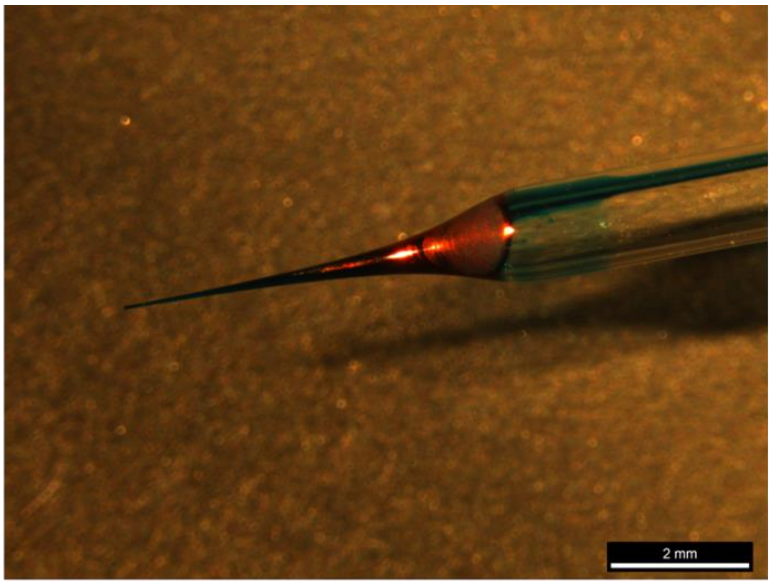


Figure 2.3: Microcapillary pipette used for injection of chick embryo.

2.23 Electroporation

The embryo was electroporated with five square waves of electric pulses at a fixed voltage of 15.5 V with a pulse length of 50 ms duration, and a gap of 100 ms. The electric field applied to the tissue caused formation of transient pores in the cells and

at the same time the negatively charged plasmid DNA was drawn towards the positive charge of the electrode and entered into the cells through the pores.

As the DNA was drawn towards the anode, the position of the electrodes in relation to the tissue altered the location of the tissue that was to be transfected (Figure 2.7). When transfecting tissue of the neural tube with DNA injected into the central canal, the electrodes were placed on either side of the neural tube (Figure 2.5). If the anode was placed or tilted ventrally to the midline, and the cathode placed dorsally, the electroporation occurred in the ventral half of the neural tube. This position was used specifically when transfection of the motor neurone cells was the target. There were two types of electrodes used: the first type for chick neural tube electroporation (this was called the 'straight wire' electrode), while the second type was used to electroporate the chick brain tissue (Figure 2. 6), this was called the paddle electrode (see Figure 2.4 for photos of both types of electrode).

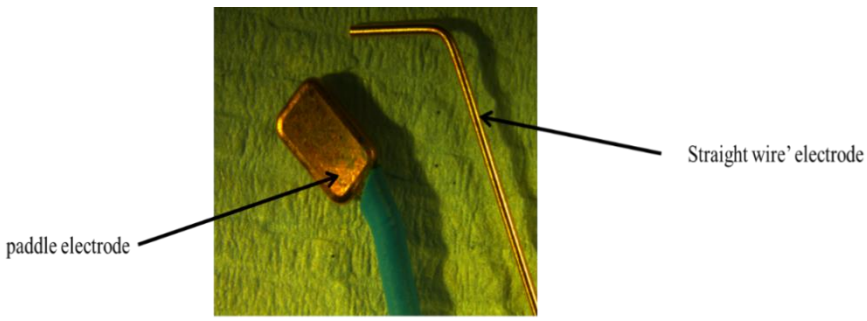


Figure 2.4: Two types of electrodes were used in chick embryo transfection. The straight wire electrode was used for transfecting the chick neural tube; the paddle electrode was used for transfecting chick brain tissue.

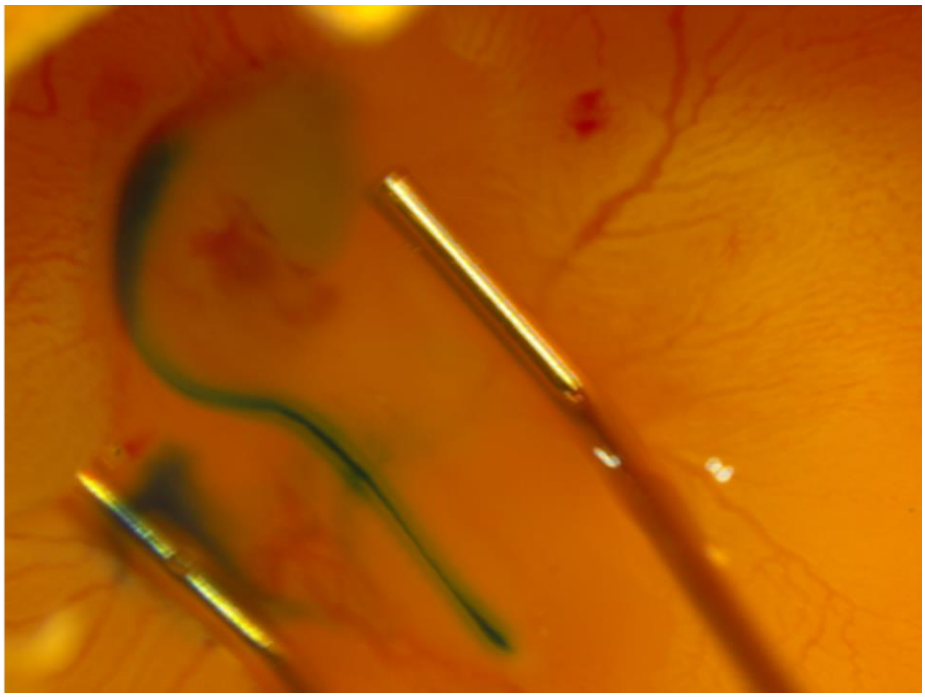


Figure 2.5: Electroporation of the chick embryo neural tube using a straight wire electrode.

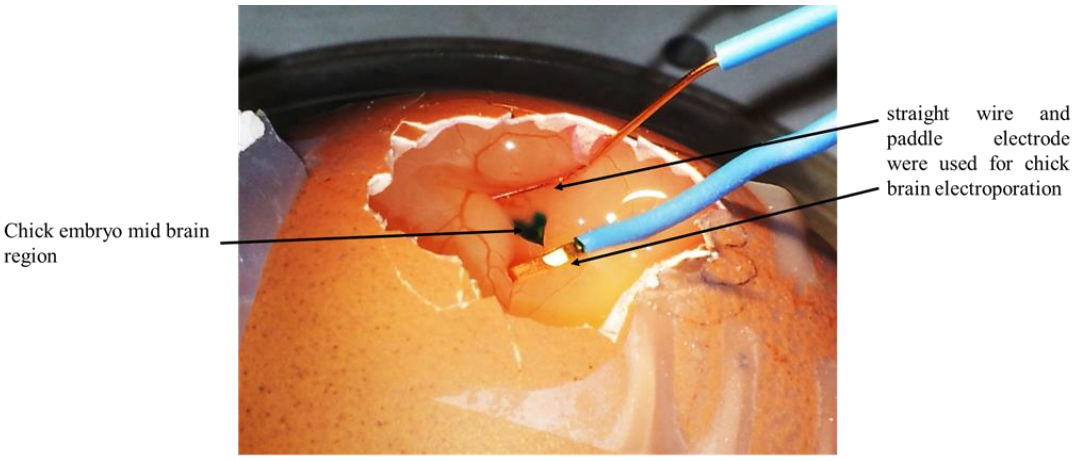


Figure 2.6: Electroporation of chick embryo mid brain region using paddle and straight wire electrodes.

Placement of the anode electrode can alter the transfection; when the anode electrode was tilted to the ventral region of the neural tube; this anode will be close to the motor column and might increase the chance of motor neurone cells being transfected (Figure 2.7).

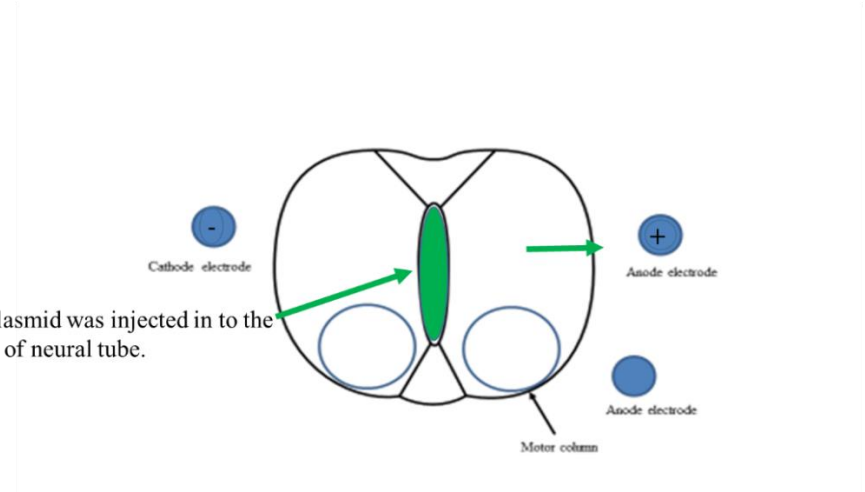


Figure 2.7: Illustration of electrode position in relation to the site of transfection of the chick embryo neural tube. Electroporation introduces the plasmid DNA to the motor neurone cells located to the anode electrode, while the other side acts as a control.

After the electroporation was completed, the window was sealed with the shell fragment, which was stuck down, or if the shell piece was missing, the window was taped over. The embryo was then allowed to develop in a humidified incubator. This specialised incubator was fitted with a specialised humidity system (Brinsea Ova Easy 380 with H22 Digital Humidity Management Module). This was used to incubate the windowed embryo. The temperature was set at 37.8 °C with the active humidity system set at 35%.

Following the incubation period, usually 48 hours after electroporation, the embryos were visualised *in ovo* under fluorescence microscope. The embryos that have expressed td Tomato fluorescence in the desired tissues, were harvested and fixed,

then visualized under a dissecting microscopy to observe the reporter gene expression GFP.

2.24 Harvesting Embryos.

The embryos were dissected out from the shell by cutting the vitelline membrane and the blood vessels from the surrounding tissue using sharp forceps. The embryo was then scooped up with curved-shank tweezers. The embryos were then washed with Hank's Balanced Salt Solution (Thermo Fisher Scientific). The embryos were then cleaned up using sharp forceps, by removing any unwanted tissue and remnants of blood vessels or vitelline membrane.

2.25 Fixing Embryos

The embryos were immersed in 4% paraformaldehyde fixative, with the fixing period being dependent on the size and age of the embryo. Typically embryos at E4-E5 would need from 1 to 2 hours to maintain the tissue integrity.

2.26 Mounting chick embryo neural tubes and cerebella on cork discs

In order to cryo-protect, the trunk of the embryo was dissected out from the whole embryo and all unwanted material was removed. The remaining tissue was submerged, until it sank, in the succession of 6%, 12% and 18% sucrose in Millonig's Phosphate Buffer (0.125 M, pH 7.4) with 0.1% NaN_3.

The required area of a cryo-protected embryo was located in the desired orientation, and mounted in a dome of optimal cutting temperature (OCT)-type embedding compound (Bright Cryo-M-Bed medium) on a cork disc.

Iso-pentane was placed in a plastic beaker which was suspended over liquid nitrogen in a Dewar's flask to cool it. The cork disc was then placed in the ultra-cold iso-pentane for about 1 min to allow the OCT and tissue to freeze, at which point it was dipped in liquid nitrogen. The tissue always stored at -80 °C

2.27 Antibody-staining of chick embryos

Sections (10 micron) of chick neural tube (E5), cerebellum (E17) were cut using the cryostat. These sections were collected and mounted on gelatine-coated slides and a circular mark was drawn with a DAKO pen around the section. Paraformaldehyde was quenched for 15 min in 0.2 M glycine pH 7.4, non-specific antibody sites blocked with diluent consisted of 1% BSA, 0.12 M (Millonig's), 0.1% Triton, 0.1% NaN_3 give for 15-30 min at room temperature. The section was stained with primary antibody at the appropriate concentration (see Table A4 Appendix) in diluent (30 min to 1 hour at room temperature), dip washed (three times in HBSS), stained with secondary antibody at 1:250 in diluent (1 hour at RT), dip washed (three times) in HBSS and once in de-ionized water and cover-slipped using a DAKO mount as adhesive.

Nuclear staining was achieved by co-application of DAPI (1:1000 from 0.1µg/ml stock) with the secondary antibody. A list of primary and secondary antibodies used in the immunofluorescence staining of the chick embryo neural tube and cerebellum is provided in Table 4 in the Appendix.

2.28 Culturing of the SK-N-AS cell line

SK-N-AS cells (human neuroblastoma cell line) were grown in media as outlined in 2.1.6 in T175 flasks. When the cells' confluency reached 70-80% they were passed into a new T175 flask. To passage the cells, the medium was removed from the flask and the cells were washed with 10 ml sterile PBS. 5 ml of 1x trypsin (Sigma) was added, washed over the cells and then removed. The flask was placed in the incubator at 37 ºC for 3-4 min until the cells started to detach from the surface. The cells were washed away from the surface flask with 10 ml of media, the cell clumps were broken up by pipetting and 1-2 ml of the cell suspension (approximately 1-2.4 million cells, depending on the cell type) was then placed in a new T175 flask with 40 ml of fresh media appropriate for the cell line. To detect infection with mycoplasma the cell lines were tested with Myco Alert mycoplasma detection kit (Lonza) every six months.

2.29 Cell counts using a haemocytometer

To determine the number of cells per ml of media a haemocytometer was used for cell counts. A T175 flask of cells at 70% confluency was passaged as stated in 2.36 up to when the cells were washed with 10 ml of media. The cover slip and haemocytometer were washed with ethanol before use. The haemocytometer has a grid which consists of 25 squares (5x5) bounded by three parallel lines each containing 25 smaller squares (5x5). The cover slip was placed over the counting surface of the haemocytometer and 20 µl of the media with cells were inserted under the cover slip. The counting surface was visualised under a light microscope on a 10x objective and the number of cells within 25 large squares bounded by three parallel lines were counted. Any cells located on the top or left hand borders of 25 squares were included in the count while any cells

on the right or bottom border were excluded. This area corresponds to 0.1 mm^3 and therefore the number of cells were multiplied by 10^4 (10000) to give the number of cells in 1 cm^3 which is equivalent to 1 ml. This gave the number of cells per ml of media.

2.30 Transfection of SK-N-AS cells

Using Turbojet transfection Reagent (ThermoScientific/Fermentas, R0531) according to the manufacturer's protocol, the cells were counted (as in 2.31) and then plated into 24-well plates at a concentration of 66,000 cells per well with 1 ml media 24 hours before transfection. 1 µg of the test reporter gene plasmid and 2 µl of Turbofect were mixed in a total volume of 100 µl of serum free media, vortexed and incubated at room temperature for 20 min. SK-N-AS cells were transfected with 100 µl of the transfection mixture, equal to 1/10 of the volume of the media. After 4 hours of incubation in 5% of CO_2 at 37 °C, the medium was exchanged with a fresh medium to reduce cell death. After 48 hours incubation the transfected cells were photographed using the inverted fluorescence microscope Leica.

Chapter 3

Evolutionary conserved regions of the human FUS gene can act as regulatory domains in the chick embryo model

3.1 Introduction

Comparative genome analyses can highlight regions of the genome that have been conserved during evolution that might play an important role in regulating gene expression. One of the previous projects in this laboratory utilised this approach to predict potential domains regulating FUS gene expression that might underlie its association as a candidate gene in ALS. The regulatory domains identified spanned the major transcriptional start site of the FUS gene and had a high degree of conservation between chimpanzees, dogs, mice and rats. These ECRs, and the putative FUS promoter, were initially selected for *in vitro* studies and subsequently analysed at embryonic day 5 (E5) of the chick embryo model *in vivo*, in order to address their potential regulatory functions in an appropriate anatomical location, specifically in the motor neurones.

3.2 Chapter Aims

- To develop a methodology using the chick embryo model to test the activity of potential regulatory elements in the human FUS gene. In this chapter, these regulatory elements are presented as evolutionary conserved regions (ECRs). ECRs were cloned into the reporter vector and transfected into the chick embryo neural tube at HH12-14 in order to assess their activity via reporter gene expression in the transfected area.

- To develop a reporter gene system using a combination of two fluorescent plasmids. The td Tomato plasmid was used to determine electroporation efficiency and a promoter-less phrGFP plasmid was used as the reporter gene. This dual plasmid system has potential advantages including that transfected embryos can be observed without embryo sacrifice, therefore enabling them to be cultured *in ovo* for 48-72 hours and thus giving a longer window of visualisation. Furthermore, this will allow gene expression to be visualised directly under fluorescent microscopy without the need for staining.

3.3 Endogenous gene expression of KIFAP3, FUS, and ELP3 in chick embryo tissue at E5.

3.3.1 Introduction

An initial experiment was conducted to confirm that KIFAP, FUS, ELP3 genes are expressed endogenously in motor neurones at this stage of development (E5). These genes were initially selected for their important association with MND, as a conducted study by (Landers et al., 2009) identified a SNPs residue within the KIFAP3 gene associated with reduced KIFAP3 expression can increase the survival rate of ALS, furthermore, ALS causal mutations were found in the FUS gene (Kwiatkowski et al., 2009), finally work by (Simpson et al., 2008) was identify ELP3 gene in ALS patients. At the early stages of this project it was planned to test putative regulatory elements in the form of evolutionary conserved regions (ECRs) from several genes known for their association with MND. When the ECR browser was applied to KIFAP3 and FUS genes, however, it showed that both genes contain evolutionary conserved regions, and these ECRs have been addressed for their activity in the *in vitro* study. The ELP3 gene, meanwhile, contains two SNPs located in the intronic region that have been associated with motor neurone disease. For this reason it was important first to confirm that KIFAP, FUS and ELP3 genes are expressed in motor neurones at this stage of development. At E5 specifically, the motor column cells in the chick neural tube are clearly apparent and developing motor neurone markers are expressed (Ericson et al., 1992).

3.3.2 Optimisation of RT-PCR assay to confirm KIFAP3, FUS, and ELP3 genes expression in chick embryo tissue at E5

The optimised RT-PCR was used to confirm the expression of genes mentioned above at E5 in the chick embryo central nerve system tissue. In brief, total RNA was prepared from the E5 chick neural tube, E15 forebrain and E17 cerebellum using the RNeasy mini kit (Qiagene). First a strand of cDNA was produced from the total RNA preparations. Platinum Taq DNA Polymerase (Promega) and primers were used for the amplification. The primers are described in detail in the table 1 Appendix. The annealing temperature was optimised while varying the cDNA concentration in the neural tube and cerebellum. The correct band sizes were amplified at the following annealing temperatures: 55 °C for KIFAP3, 57 °C for FUS and 59 °C for ELP3. The PCR products were visualised by agarose gel electrophoresis using a 1 kb fragment of KIFAP (Figure 3.1), a 994 bp fragment of FUS (Figure 3.2) and a 997 bp fragment of ELP3 (Figure 3.2).

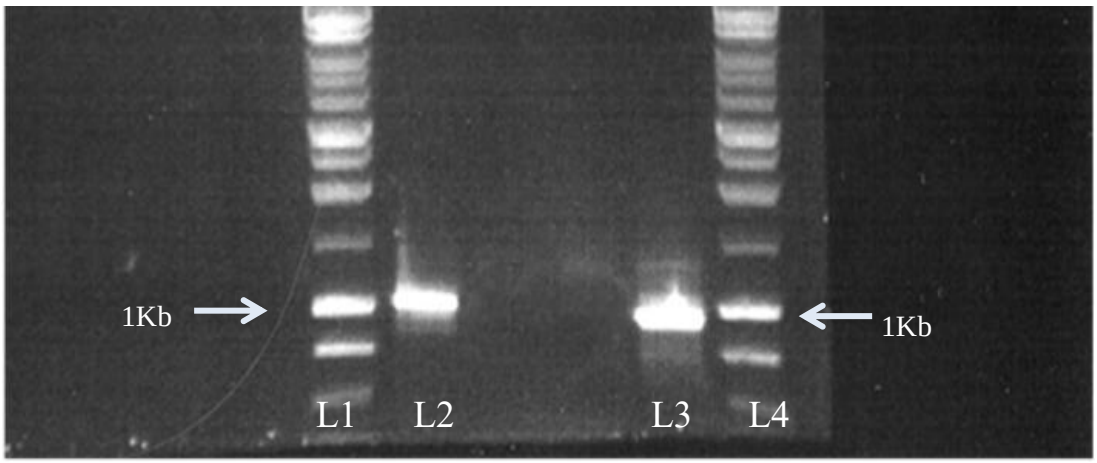


Figure 3.1: Endogenous expression of chick KIFAP3 gene. Expression was detected in the chick CNS. Total RNA was purified from the E5 neural tube, E15 forebrain, reverse transcribed to cDNA and then used as a PCR template.L1 1kb ladder, L2 KIFAP3 in neural tube, L3 KIFAP3 in forebrain, L4 1kb ladder.

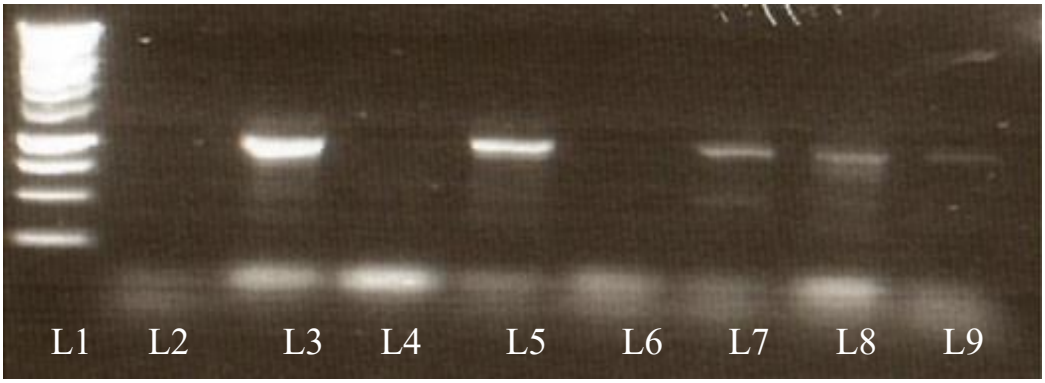


Figure 3.2: Endogenous expression of FUS and ELP3 genes in the chick CNS. Total RNA was purified from E5 neural tube, E17 cerebellum, reverse transcribed to cDNA and then used as a PCR template. L1 1 kb ladder, L2 negative control, L3 FUS in neural tube 2 µl, L4 FUS in neural tube 1 µl, L5 FUS in Cerebellum, L6 negative control, L7 ELP3 in neural tube 2 µl, L8 ELP3 in neural tube 1 µl, L9 ELP3 in cerebellum.

3.3.3 Selection of FUS gene in an analysis of potential regulatory domains that regulate gene expression

Putative regulatory elements in the form of ECRs from KIFAP3, FUS and ELP3 were tested *in vitro*. The activity of KIFAP3 and ELP3 genes were not strongly supported *in vitro,* but a positive result was achieved from the evaluation of FUS gene ECRs *in vitro* (Khursheed et al., 2015). For this reason it was decided to concentrate on FUS gene analysis both *in vivo* and *in vitro* and not to continue the initial work with KIFAP3 and ELP3 (data not shown).

3.3.4 Injection and transfection of the chick embryo neural tube with the GFP reporter vector alone

Many elements of *in vivo* injection and the electroporation technique were developed and refined during this project in order to obtain the best results. This included egg incubation, embryo maintenance, and modification of injection and electroporation parameters such as the microinjection technique and electrode positioning.

Initial experiments were conducted to assess the GFP-reporter vector expression in the chick embryo neural tube. Although previously conducted in the lab, these experiments are technically demanding. For this reason, trial transfections were conducted to ensure consistency with data obtained from previous similar studies in the lab. This involved *in vivo* injection to determine the electroporation efficiency by demonstrating the expression of IRES-GFP (Figure 6.8 Appendix) driven under a constitutive promoter in the chick embryo (Yaneza et al., 2002). Injection mixtures consisted of Fast Green dye dissolved in PBS to allow visualisation of the injection site under bright-field microscopy, mixed with the GFP-reporter vector. The lumen of the chick embryo neural tube at HH12-14 was microinjected with the mixture and two

3 mm electrodes were placed over the neural tube. For electroporation, 5 square-wave electric pulses were applied at a fixed voltage of 15.5V with a pulse length of 50 ms duration and a gap of 100 ms. Determining the optimum electrode position in relation to the injected area was of paramount importance, but this was difficult because of the instability of the embryo and its environment. Achieving good electrical contact with the egg yolk was essential and this was ensured by removing the vitelline membrane and placing a few drops of PBS over the embryo. It was found that the electrodes needed to be positioned such that the injected area was centred between the two electrodes in a parallel orientation. To ensure that the embryo was not disturbed, and to avoid bleeding through contact with the major blood vessels, this process was carried out with extra diligence. Optimal electroporation was confirmed by detecting the expression of GFP in the neural tube under fluorescent microscopy.

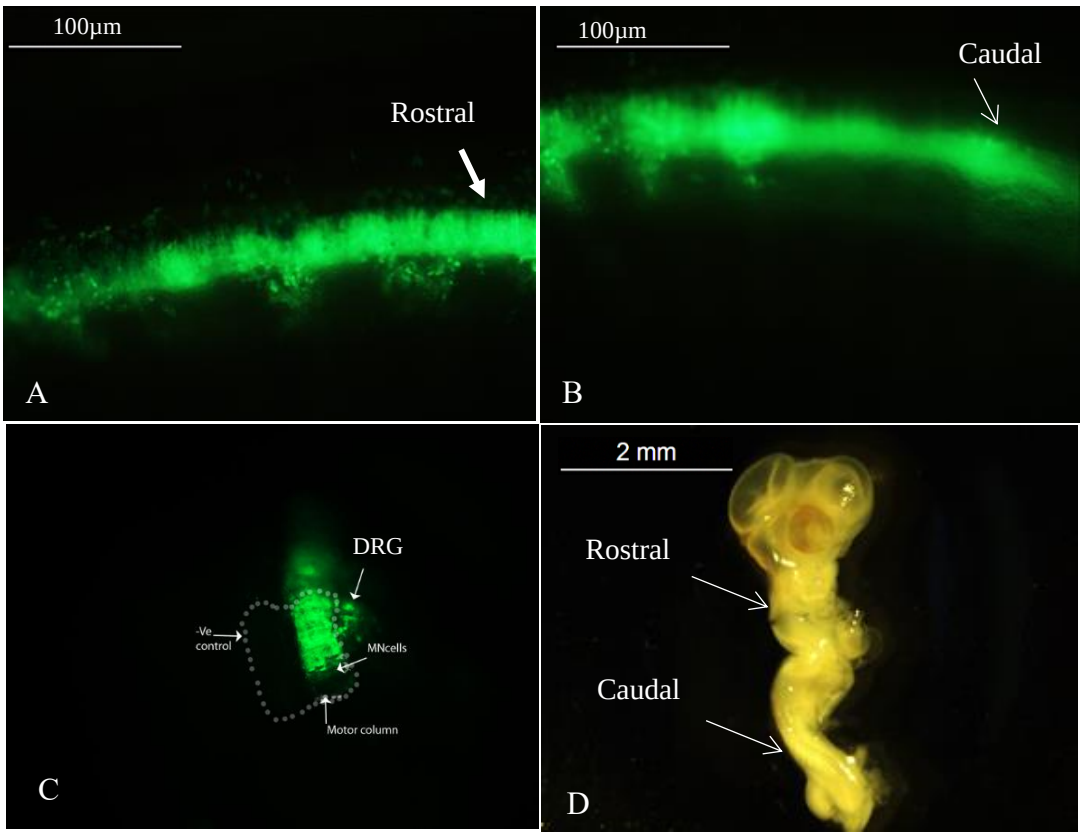


Figure 3.3: Representative images of Chick embryo transfected with GFP. Whole mount (A, B) and cross-section of the dissected neural tube using a lab blade (C) of embryonic day 5 chick showing fluorescence (A,B) indicating the location of electroporation in the rostral region (A) and caudal region (B) of the neural tube following successful micro-injection and electroporation of the neural tube. (C) Showing fluorescence in dorsal root ganglia, motor neuron cells and motor column; (D) whole mount chick embryo under bright field microscopy. Images are representative of 3 similar experiments using 10 embryos.

As shown in Figure 3.3 this vector drives the extensive expression of GFP in the rostral (A) and caudal (B) regions of the neural tube. A cross-section of the neural tube (C) shows that most of cells located proximally to the positive electrode expressed detectable GFP, whereas no expression was detected on the contralateral side, meaning it could act as a negative control. Strong GFP expression was detected in >90% of the samples at 48 hours, indicating the effectiveness of this method for introducing exogenous DNA. The bright field microscopy image (D) illustrates the rostral and caudal regions of the chick embryo neural tube at E5. This experiment established that the proposed electroporation system works well when GFP plasmid is transfected at HH12 and analysed at E5.

3.3.5 Chick embryo neural tube at E2-3 was transfected with Hb9 enhancer as a positive control

The electroporation technique was now used to transfect the motor neurone cells from the chick embryo neural tube at HH 12-14. For this, the p1229 vector reporter gene system was used (a kind gift from Dr Michael Lyons) (Figure 6.7 Appendix), which contains the LacZ coding sequence driven by the human β-globin minimal promoter. An Hb9 enhancer sequence from the HOXb9 gene was cloned in the p1229 vector to confer detectable expression in motor neurone cells (Nakano et al., 2005). The transcription factor binding to the Hb9 enhancer activates the human β-globin minimal promoter driving LacZ gene expression and resulting in the expression of β-galactosidase (β-gal) in the cell. β-gal is therefore expressed only where the Hb9 enhancer is active. Hb9 is also only expressed in developing spinal motor neurones and is considered to be a strong marker for motor neurone cells (Tanabe et al., 1998, Arber et al., 1999).

A mixture of 1 µg/µl of GFP plasmid and 2 µg/µl of Hb9 enhancer plasmid with 1 µl of Fast Green were dissolved in PBS and microinjected into the neural tube of chick embryos. 14 embryos were successfully injected and electroporated. Following incubation for 48 hours at 37 °C three of these embryos showed strong GFP expression in the neural tube (Figure 3. 4, B) and were processed as described in the material and methods chapter embryo staining section, remaining embryos has GFP expression in superficial tissues that allocated around the neural tube were excluded from the experiments as the transfected tissue has no motor neuron cells After incubation for a further five hours at 37 ºC all three embryos showed a blue staining in the neural tube in, co-localised with the observed GFP fluorescence (Figure 3.4 A).

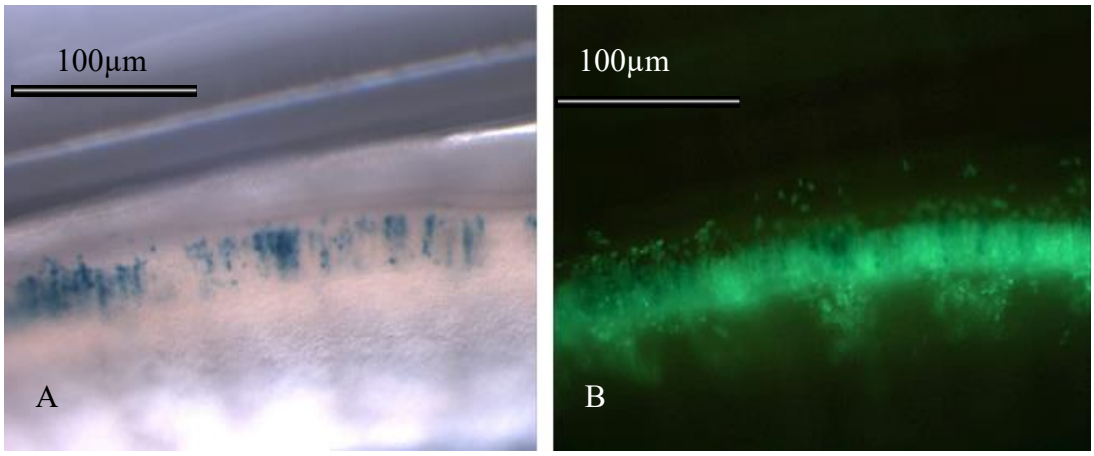


Figure 3.4: Representative images of Chick embryo transfected with Hb9 enhancer. Whole mount (A,B) dorsal view showing the Hb9 enhancer LacZ-driven blue staining in the spinal motor neurons of a chick embryo at E5 (A) GFP (B) following successful micro-injection and electroporation Scale bar 100 µm. Representative of 3 similar experiments.

The chromogenic signal of β-gal expression can only be visualised following sacrifice of the embryo. For β-gal stain development the tissue must be incubated in X-gal staining solution for 5 hours at 37 ˚C. This process is time-consuming, however, making β-gal reporters a less than ideal choice for expression studies in the chicken neural tube.

3.3.6 Development of dual plasmid system using a combination of two fluorescent plasmids

Rather than use the p1229 Hb9 enhancer as a reporter gene system, therefore, a new reporter gene system was introduced. This new system used a combination of two plasmids, both of which expressed fluorescent proteins so that the transfected embryo could be monitored and observed using florescent microscopy *in vivo.* This new system consisted of a plasmid expressing td Tomato under the control of a constitutively active chick β-actin promoter and shows strong red fluorescent illumination when observed under a fluorescent microscope with a red filter. td

Tomato plasmid has td Tomato sequences and is genetically modified from *Discosoma striata* (DsRed). It has a bright fluorescent protein (FP), and has tandem repeat dimer FP. This was created by coupling two td Tomato units with a 12 amino acid linker. td Tomato has many advantages, including compatibility with the existing lab microscope and Leica red filter set. It also has a fluorescence excitation spectrum peak at 554 nanometres and an emission spectrum peak at 581 nanometres along with a high photostability region. td Tomato was constructed was used as a marker protein to indicate successful microinjection and electroporation (Strack et al., 2009) . With td Tomato plasmid a phrGFP plasmid was used as an alternative to the p1229 Hb9 enhancer and acted as a reporter gene plasmid. This dual plasmid system has many advantages, including the fact that transfected embryos could be observed without embryo sacrifice and, therefore, can be cultured *in ovo* for 48-72 hrs giving a longer window of visualisation. Furthermore, gene expression can be directly visualised under fluorescent microscopy without the need for staining.

3.3.7 Testing of the fluorescent overlap between the GFP channel and red channel

A set of three experiments were conducted to observe the fluorescent properties of GFP as a reporter and td Tomato as a marker (Figure 6.9 Appendix) and to confirm that there were no potential artefacts or fluorescence overlaps between the green and red channels when visualising plasmid expression. To test fluorescence overlap or artefacts related to the td Tomato plasmid, an injection mixture containing Fast Green dye in PBS and td Tomato plasmid was introduced into the chick embryo neural tube at development stage HH 12-14. When visualised under the red filter, a strong red fluorescence was observed in the neural tube of successfully electroporated embryos. When samples were observed under fluorescence it is very important to avoid the loss of signal through photo-bleaching. In my study it didn't seem to affect the tested samples under the conditions I followed. Photo-bleaching was avoided by reducing the illumination intensity, and by the use of robust fluorophores. Given these results, the td Tomato plasmid could act as suitable marker in chick embryo tissue with sustained florescent intensity maintained for >48 hours, and even up to E5 (Figure 3.5A). Moreover, when the same image was viewed under the green filter there was no fluorescence, indicating an absence of fluorescent overlap or artefacts (Figure 3.5C).

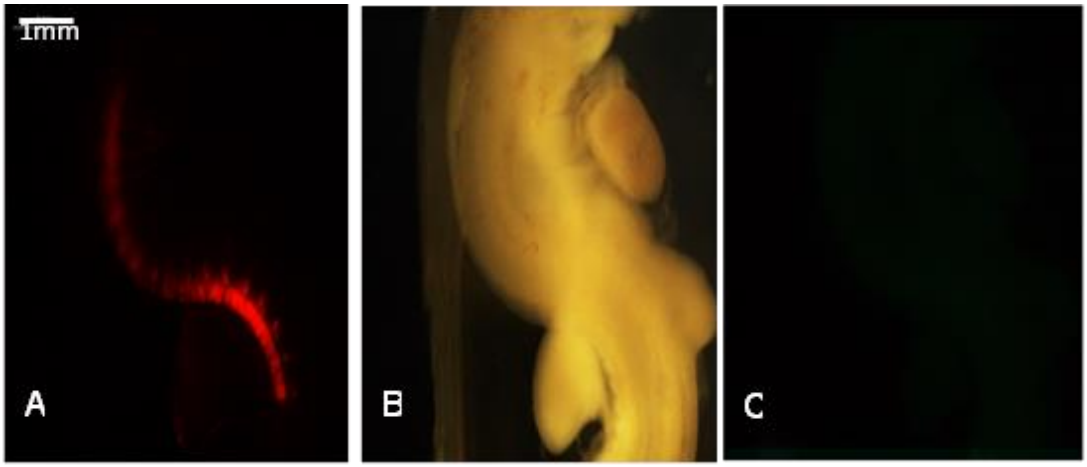


Figure 3.5: Representative image of Chick embryo transfected with td Tomato plasmid in the neural tube. Chick trunk (A, B) At E5 showing fluorescence (A) the expression of the tdTomato is detected in the neural tube by a red filter indicating the location of the electroporation. td Tomato has expressed under the beta actin promoter and has a strong red fluorescence, therefore needing short exposure times, usually around 500 ms (B) The same image under bright field microscopy (C) No fluorescence under green filter.

To test for fluorescence overflow or artefacts when the IRES GFP plasmid alone was electroporated into the chick embryo neural tube, the IRES GFP has a major excitation peak at 488 nm, and peak emission at 509 nm. An electroporated embryo was visualised under the green filter and a strong green fluorescence was apparent in the transfected area, whereas when viewed under the red filter, no fluorescence was observed, indicating that IRES GFP could be used as a marker plasmid without any fluorescent overlap (Figure 3.6).

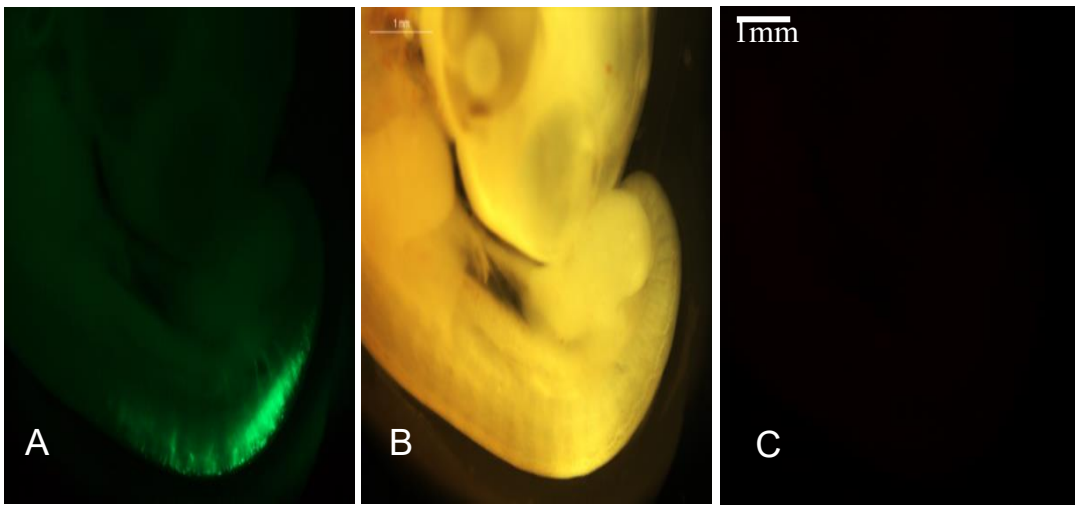


Figure 3.6: Image of Chick embryo transfected with GFP in the neural tube. Whole mount (A, B, C) of embryonic day 5 chick showing fluorescence; (A) the expression of GFP is detected in the neural tube by green filter indicating the location of the electroporation. GFP was detected with long exposure times usually between 1-1.5 s. (B) Same image under bright field microscopy (C) with no fluorescence under a red filter (D) Schematic representation of GFP reporter gene vector.

Finally, the td Tomato and IRES GFP plasmids were tested together in chick embryos. Both exhibited strong fluorescence under their corresponding filter. The transfected embryo in (Figure 3.7) shows td Tomato expression in the neural tube (A) and GFP expression in (C). These findings suggest that both plasmids could be co-injected for subsequent visualisation experiments; that both exhibited strong expression in the appropriate anatomical location, and that each had independent intense signals.

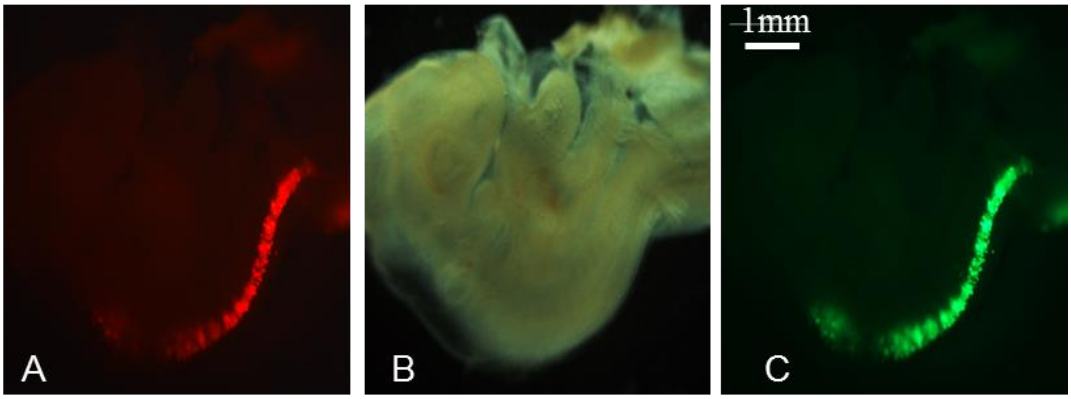


Figure 3.7: Chick embryo transfected in the neural tube with both td Tomato and GFP. Whole mount (A, B, C) of embryonic day 5 chick showing fluorescence (A, C) The expression of td Tomato and GFP are detected in the neural tube; the former by a red filter and the latter by a green filter, indicating the location of the electroporation. (B) The same image under bright field microscopy.

3.3.8 Promoter-less phrGFP cannot drive GFP expression without a suitable promoter

When the regulatory domains from the human FUS gene were observed, it was decided to use another type of GFP as a reporter gene. This type was called humanised recombinant GFP (hrGFP). It has a humanised recombinant GFP gene and a SV40 polyadenylation signal. This version does not have a promoter and thus provides an ideal option to clone the promoter/enhancer of interest in extensive multiple cloning sites (Figure 6.10 Appendix). In addition, it has a major excitation peak at 500 nm and an emission peak at 506 nm. This vector has been tested in different cell lines and shows a high level of expression of functional fluorescent protein, and is more consistent, less toxic than EGFP, and the GFP produced from transfected cell lines is more efficient (Dardalhon et al., 1999). ECRs from the FUS gene were cloned into phrGFP so that the regulatory activity of these FUS ECRs could subsequently be tested on our own minimal promoter. To achieve this, the phrGFP was first tested alone to confirm that this plasmid cannot drive GFP expression without a suitable promoter. Chick embryos at HH12-14 were microinjected with phrGFP and td Tomato plasmids and these were electroporated into the neural tube. 48 hours post transfection; successful embryos exhibited td Tomato expression under a red filter, but did not show any GFP expression under a GFP filter, confirming that in the absence of a promoter, the phrGFP alone cannot drive GFP expression in the chick embryo neural tube. td Tomato was included as an internal control to ensure successful electroporation (Figure 3.8).

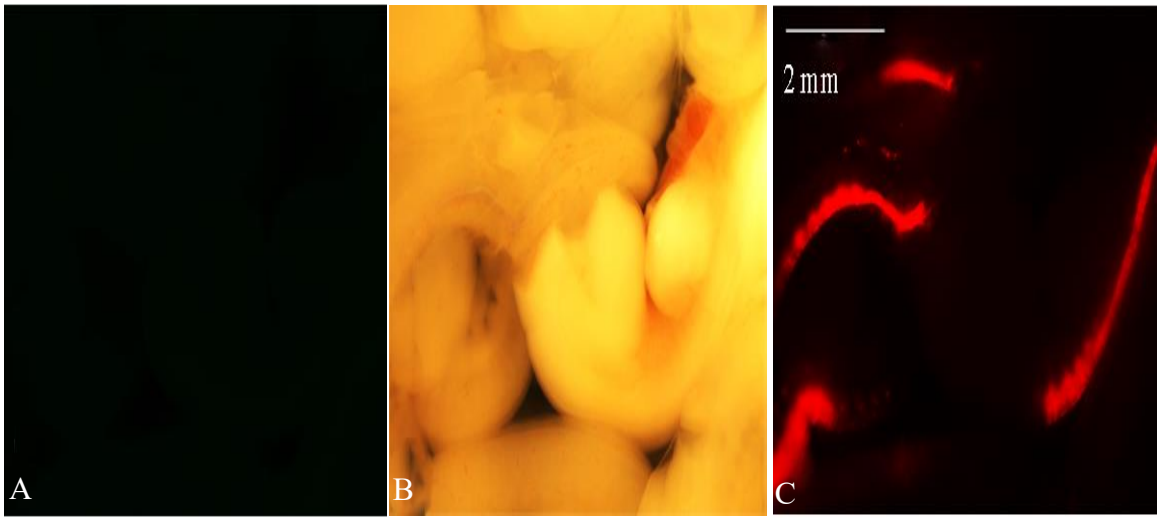


Figure 3.8: Chick embryos transfected in the neural tube with phrGFP plasmid with absent promoter and presence of td Tomato. Trunks of chick embryos at embryonic day 5 showing no GFP fluorescence (A) no expression of the GFP is detected in the neural tube by green filter indicating the phrGFP with no promoter has no ability to drive expression. (B) The same image under bright field microscopy (C) td Tomato fluorescence under a red filter.

3.3.9 The FUS gene was expressed ubiquitously in the chick E5 neural tube and E17 cerebellum

To establish the validity of analysing reporter gene expression in embryonic day 5 (E5) embryos, it was important to validate endogenous chick FUS expression in the chick embryo neural tube at E5. Endogenous of FUS expression was detected previously in the neural tube and forebrain by RT-PCR, thus this experiments was done by immunohistochemistry to confirm FUS gene expression in the cells of neural tube and specifically in the motor neurone cells. Cross-sections of the neural tube at E5 and the cerebellum at E18 were stained with an anti-FUS antibody (Abcam ab 23439). Neural tube stained sections demonstrated that FUS had ubiquitous expression in the embryo at these stages (Figure 3.9 A). Cell nuclei were detected by DAPI staining in Figure 3.9 B and D. The E18 cerebellum cross-section also exhibited ubiquitous FUS expression in cells of all regions including cells of the molecular layer, Purkinje, and granular layers (Fig. 3.9 C). The negative control, lacking a primary antibody, did not show any staining (Fig. 3.9 E).

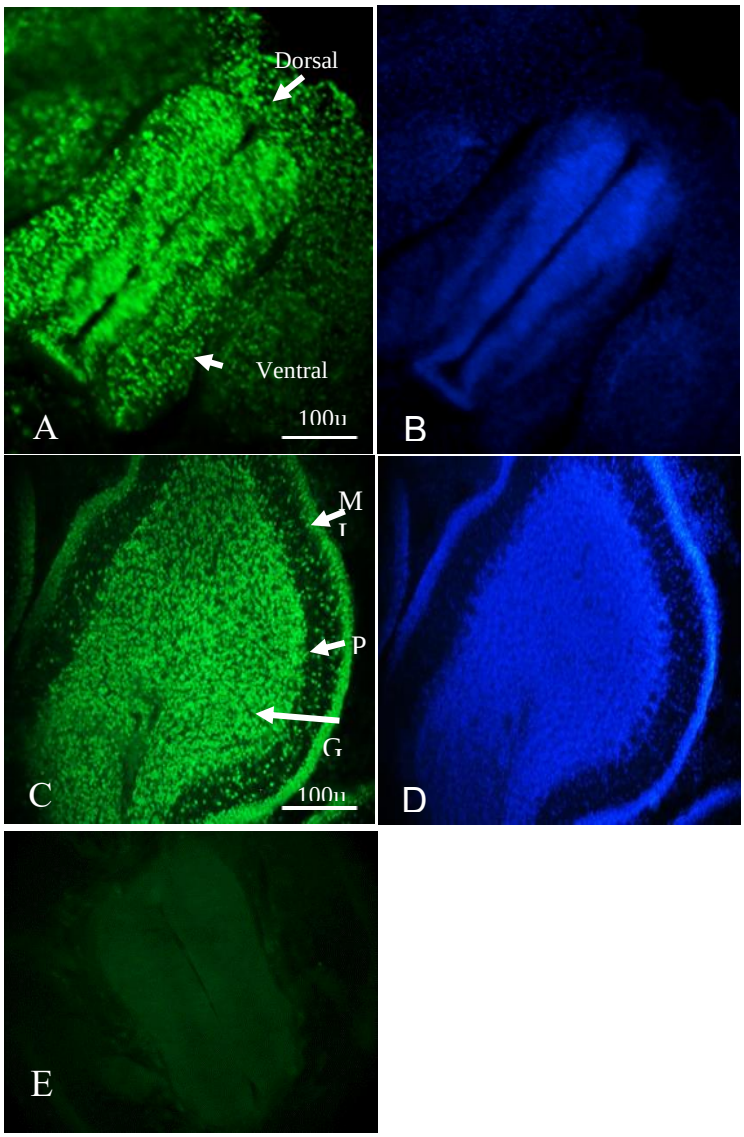


Figure 3.9: FUS is expressed in the neural tube of E5, and in the cerebellum of E18 chick embryos. Frozen section of chick embryo E5, E18 stained with (A, C) FUS antisera, (B, D). DAPI. FUS is ubiquitously expressed in all cells in the section, including cells in the developing motor column and all cells in the cerebellum section molecular layer, Purkinje layer (PL), granular layer (GL). (E) No primary. Scale bar 100 µm (A, C).

3.3.10 The location of motor neurone cells was identified in the chick neural tube by Hb9 antibody staining

Prior to testing the regulatory propriety of FUS ECR in the chick embryo it was important to identify the region of the developing motor column and motor neurones in the chick embryo neural tube, so as to address the location and site that should be transfected with FUS regulatory domains. To achieve this aim immunohistochemically staining of frozen sections of E5 chick embryo neural tube was carried out using a Hb9 antibody (obtained from the University of Iowa) (Nakano et al., 2005, Ovcharenko et al., 2004); a secondary antibody was used –goat anti-mouse Alexa 488 (Invitrogen). The stained section demonstrated Hb9 expression specifically in the nuclei of motor neurone cells (Figure 3.10 A). The negative control, lacking a primary antibody, did not show any staining (Fig. 3.10 B).

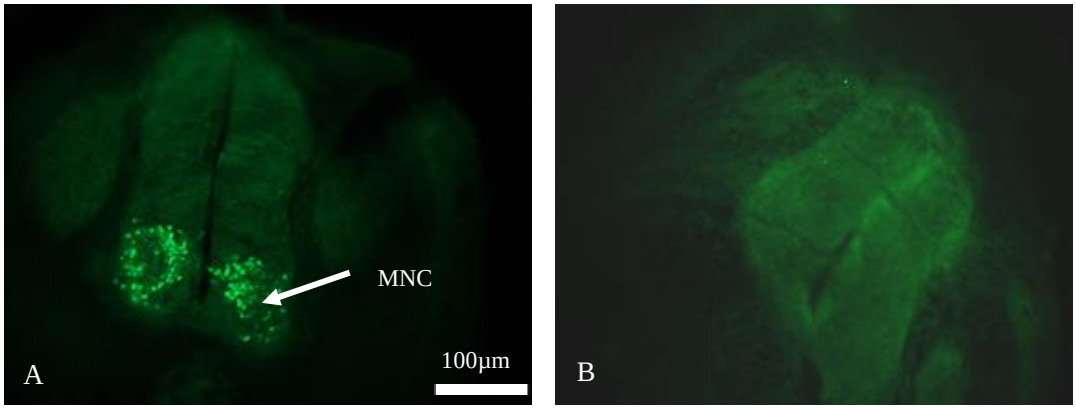


Figure 3.10: The region of the developing motor column was identified by the Hox gene family. (A) A cross-section of chick embryos unlabelled neural tube at E5 stained for a motor neuron marker protein HB9. The white arrow indicates motor neuron cells (MNC). (B) No primary. Scale bar 100 µm.

3.3.11 An upstream (ECR) and intronic (INT) region are conserved during evolution

In silico and functional *in vitro* analysis of the potential regulatory elements of the FUS gene was undertaken by my colleague Dr Thomas Wilm at the University of Liverpool. This data has been presented in a published paper (Khursheed et al., 2015). The data from the *in vitro* experiments has been summarised here to show the functional properties of the ECRs. Comparative genome analysis was conducted using the UCSC genome browser and the ECR browser (Ovcharenko et al., 2004, Kent et al., 2002).

Comparisons of genomic sequences between distant organisms using UCSC genomic and ECR browsers can highlight candidate regions of the genome that have been conserved over a long period. This analysis can identify noncoding elements that may have critical biological functions. When this analysis was applied to the FUS gene a high degree of conservation was identified both 5' and 3' of the predicted major transcription start site (MTSS), located at Chr16: 31,191,429 (Figure 3.12) The highlighted area showed more than 70% conservation between human, chimpanzee, dog, and mouse species, but this region was not conserved in the chicks and the peak height corresponded to the percentage conservation with over 100 bp regions. The width of the peak represented the number of bases in the domain.

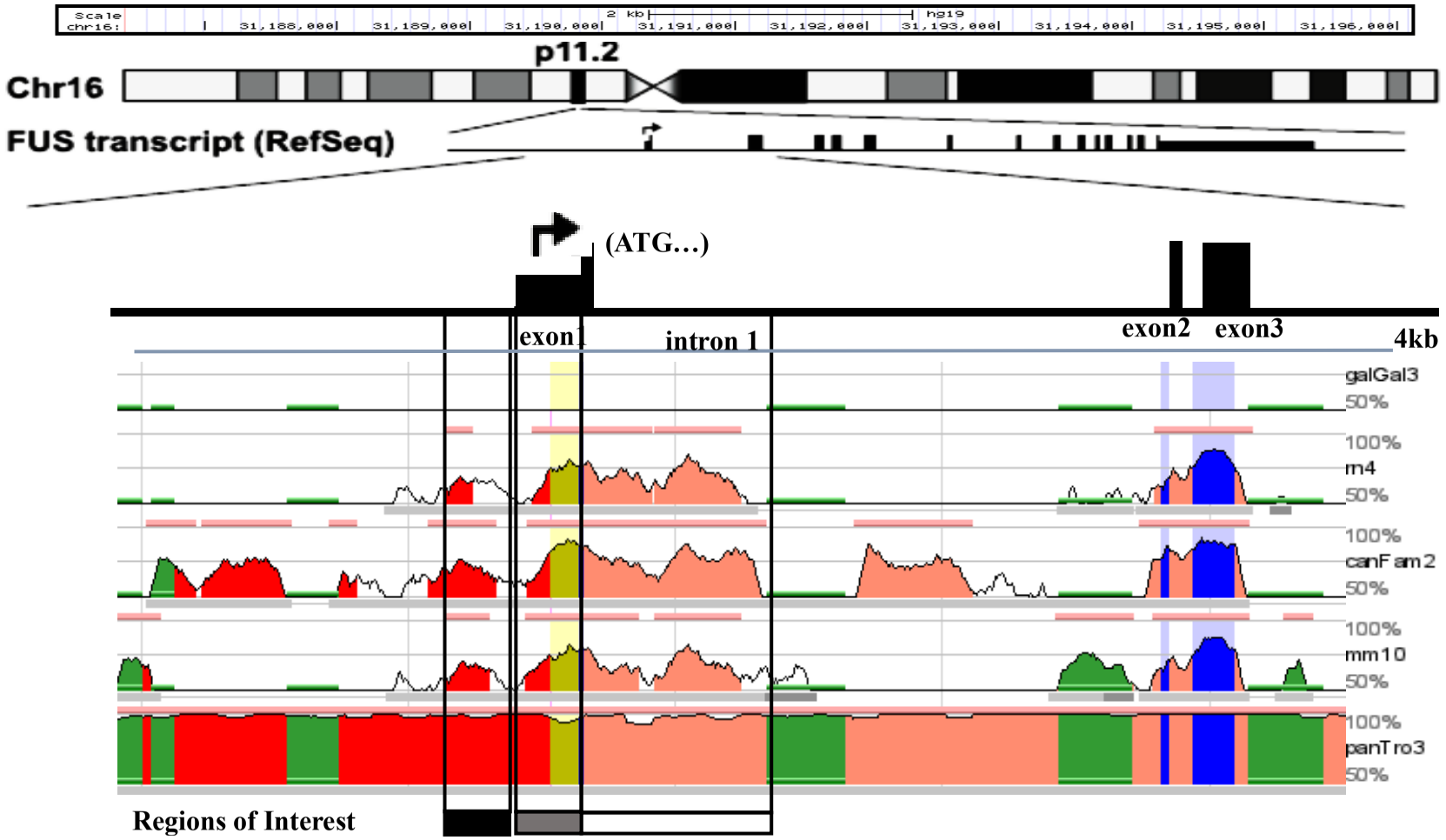
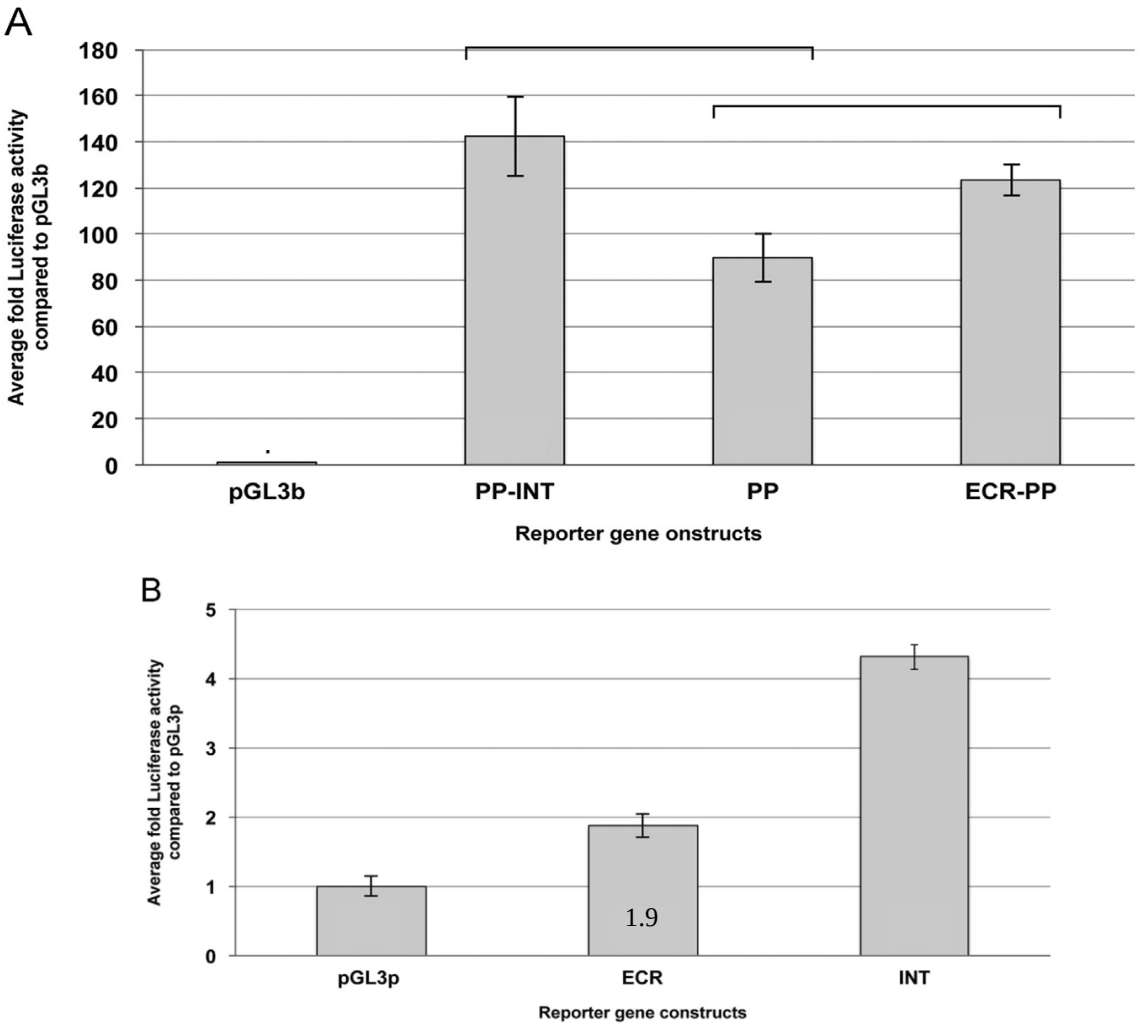


Figure 3.11: Schematic representation of the human FUS gene locus. The human FUS gene is located in the position p11.2 on chromosome 16. The above figure shows a comparison analysis of the human FUS gene aligned with corresponding FUS genomic regions of chimpanzee, dog, and mouse and rat species conducted using the ECR browser (Ovcharenko et al., 2004). This comparison highlights the important ECRs located in the coding and noncoding sequences within the FUS gene locus. The major transcriptional start site is represented by the bent arrow. The ATG indicates the first translated codon. Three important regions were identified including non-coding ECRs, the proximal promoter (PP), intron 1 (INT). Region coloured in red correspond to intergenic regions, Green: transposons and simple repeats, while yellow boxes correspond to UTRs, sequences are highlighted in salmon they are intronic region and blue boxes are belong to coding exons.

Firs predicted domain identified was termed the proximal promoter (PP) a predicted domain continuous with MTSS. This region contained 137 bp upstream of the MTSS and 106 bp of untranslated parts of exon 1. The FUS promoter has no TATA box sequence but does have stretches of a G C rich region, and has several transcription binding sites for transcription factors AP2, Sp1 and GCF (Åman et al., 1996). The FUS promoter is involved in the initiation and or elongation of transcription by interacting with the RNA polymerase II and the TFIID complex (Bertolotti et al., 1998). Additional conserved region was observed; further upstream of the proximal promoter 311 bp sequences in length and located 29 bp 5' to PP; co-ordinates Ch16:31,190,954 -31,191;264. The last identified region was observed at 5' portion of the first intron continuous with exon 1 this domain termed as INT (704 bp). This showed the most conserved region between compared species and thus potentially represents an important regulatory domain (Figure 3.11).

Now that these key regions have been identified, they will be cloned into the expressions systems that have been developed and verified so far in the thesis, and then evaluated to see if they have regulatory functions or not. These domains are shown to have activity *in vitro,* and their regulative function in both neuroblastoma and the hybrid neuroblastoma/motor neurone cell line NSC-34 cells is briefly demonstrated (paper ref). First, constructs containing ECR and INT were both maintained contiguously with the predicted MTSS of FUS and they are in the same order as they appear in the genome. These constructs were tested in the neuroblastoma cell line, SK-N-AS. The *in vitro* experiments were undertaken in the lab by Dr Thomas Wilm and compiled into a published (Khursheed et al., 2015). The data of the *in vitro* experiments has been summarised here in order to demonstrate the regulation activity of FUS domain. Previously, SK-N-AS had been confirmed as exhibiting endogenous

FUS gene expression. Construct containing PP (243 bp) was shown to act as strong promoter and drove the expression of the luciferase marker gene when it tested in the promoter-less pGL3b vector (p< 0. 002). Constructs that contained INT PP, or ECR PP are shown to act as regulatory domains (Figure 3.12A). When they cloned up a stream of heterologous SV40 promoters in the pGL3p reporter gene vector, these regions showed themselves to be positive regulators of expression, with INT having a significantly greater effect in this respect (Figure3.12B). When both domains were tested in the hybrid neuroblastoma/motor neurone cell line (NSC-34 cell with pGL3) they were both shown as being able to act as positive regulators. Furthermore, the INT domain showed a greater activity than that observed in SK-NA-S, while the ECR was shown to have the same level of activity (Figure 3.12C).



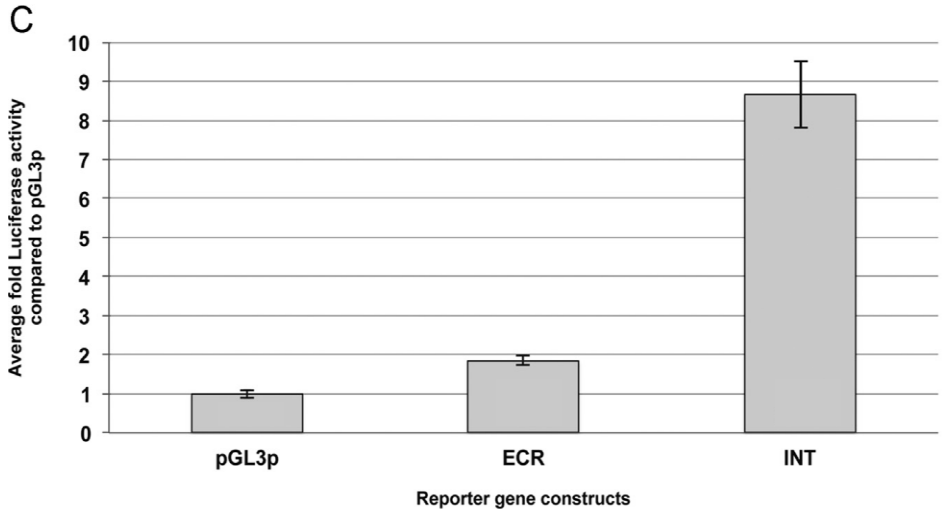


Figure 3.12: Demonstration of FUS reporter constructs' activity in SK-N-AS and NSC-34 cells. (A) FUS promoter constructs were cloned in the promoter-less pGL3b and tested in SK-N-AS cells. The FUS PP drove strong expression of the luciferase marker gene ($p<0.002$) while constructs with either ECR or INT in combination with the PP increased levels of expression and showed higher levels of transcription activity ($p<0.04$, $p<0.03$ respectively) than PP alone. (B, C) the upstream ECR and intron INT from the FUS regulatory domain when cloned into the SV40 promoter pGL3p and expressed in SK-N-AS cells (B) and NSC-34 (C) increased levels of luciferase activity compared to the SV40 promoter in SK-N-AS cells (p,0.02, $p<0.004$). Bars represent the fold change in firefly luciferase activity normalized against *renilla* luciferase activity, and means were calculated from a number of experiments. Transfection experiments were performed in double triplicates taken from (Khursheed et al., 2015)

3.3.12 FUS regulatory domains enhance expression in a chick embryo *in vivo* model

To test the functional activity of the FUS regulatory domains with their own proximal promoter in the chick embryo neural tube, FUS fragments PP (243 bp) and ECR PP (582 bp) were excised from their original vectors using *sacI* and *bglII* and subsequently cloned into the *Bam* HI/*SacI* sites of the MCS of the promoter-less

phrGFP vector (Stratagene, UK) upstream of the GFP reporter gene. The constructions of both plasmids were described in detail in Chapter 2 section 2.15.15.1 and were a kind donation from Dr Vivien J. Bubb University of Liverpool). The resulting constructs were named FUS PP phrGFP and FUS ECR PP phrGFP.

3.3.13 Construction of INT PP phrGFP and INT ECR PP phrGFP reporter gene vectors

To generate the above plasmids the INT1 704 bp fragment was PCR-amplified using proofreading Phusion DNA polymerase (NEB, UK); the primers and PCR reaction conditions used were described in Chapter 2 section 2.15.15.2 The PCR reaction product was visualised by 1.2% agarose gel electrophoresis (Figure, 3.13 lane 2).

The band at 704 bp was excised, purified, and digested with *BglII* and *Nhel.* This fragment was subsequently ligated into FUS PP phrGFP and FUS ECR PP phrGFP vectors. The resulting vectors were transformed, and minipreps analysed by double digestion with restriction enzymes/endonucleases *BglII* and *EcoRV* resulting in partially cut products. The digested minipreps were screened on a 0.8% agarose gel. If the desired plasmid was successfully constructed, it would give expected fragment sizes of 1300 bp and 950 bp (Figures 3.14 and 3.15). Clones containing the expected inserts were expanded overnight, purified and sequenced for perfect homology and then concentrated for use in the injection mixture.

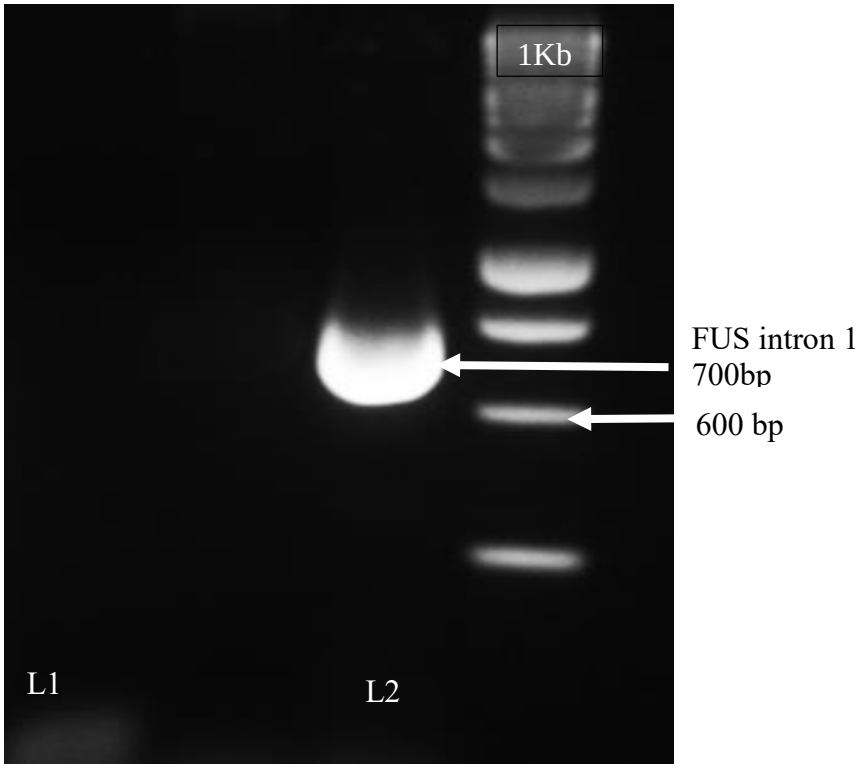


Figure 3.13: 1.2% agarose gel showing FUS intron1 amplicon. L1 is the negative control; L2 is, approximately, the expected 700bp FUS intron1 L3 1kb ladder.

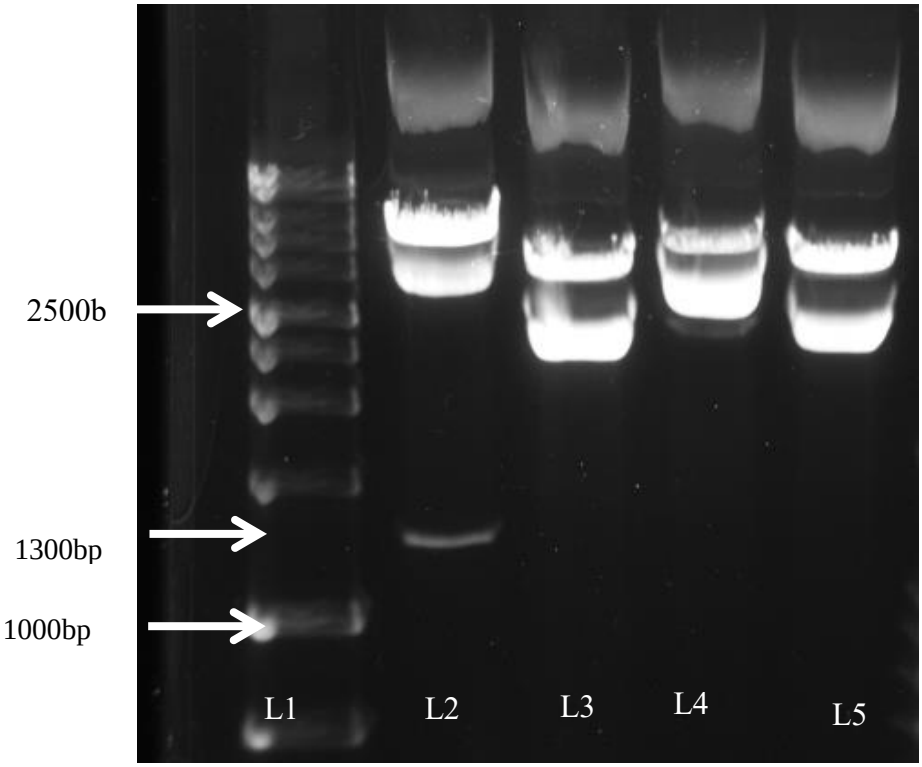


Figure 3.14: The 1300 bp band size of INT ECR PP. Double digestion of FUS INT ECR PP from the phrGFP vector with *Eco*RV and *Bgl*ll restriction enzymes. Vector of phrGFP 3.7 kb L1 1kb ladder, L2 digested colony with the insert, L3, L4, L5 are empty colonies.

Insert fragment of 950 bp

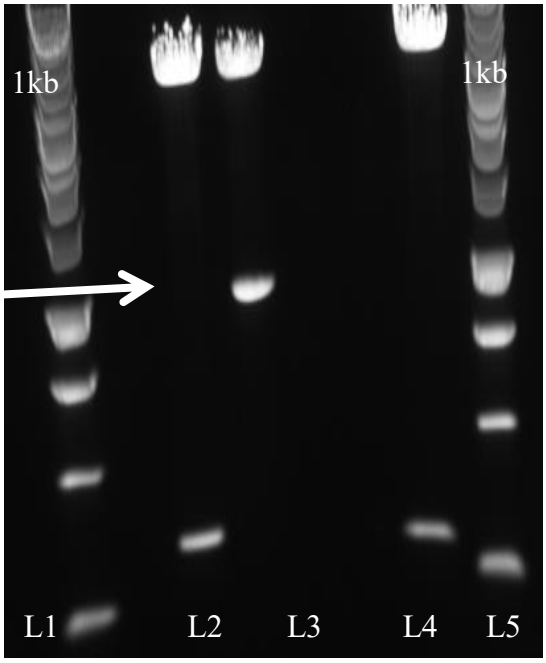


Figure 3.15: The 950 bp band size of INT PP. Double digested of INT PP from phrGFP with *Bglll*, *Eco*RV Vector of phrGFP 3. 700 bp L1 1kb ladder, L2 empty colony, L3 digested colony with insert, L4 empty colony, L5 1kb ladder

3.3.14 Demonstration of the FUS regulatory domains' activity in the chick embryo neural tube at E5

The potential activity of the predicted FUS gene regulatory regions was assessed in the cell model and the data generated from FUS domains activity show to act as activators of the expression. The order of the regulatory regions where switched to enable the activity of each region to be tested individually. The results were able to demonstrate the activities of the FUS regulatory domains in the targeted anatomical location, specifically in the motor neurones of the chick embryo neural tube. The FUS regulatory domains that were tested in chick embryos were (FUS PP, FUS ECR PP, FUS INT PP, and FUS INT ECR PP Figure 6.11, 6.12, 6.13, 6.14 Appendix) in each transfection the total DNA of the FUS constructs were maintained to the recommendation range that adopt from Cold Spring Harbor Manual (Sambrook et al., 1989).

3.3.15 FUS ECR PP phrGFP can act as activator in the chick embryo neural tube

To test the activity from FUS ECR PP phrGFP plasmid in the chick embryo (Figure 3.16J), the lumen of the developing chick embryo neural tube at HH 12-14 was injected with 1µl each of FUS ECR PP phrGFP plasmid at (3.6 µg), of tomato plasmid (1 µg/µl), 1µl PBS, and of Fast Green (1 µg/µl) and then transfected by electroporation. In each transfection the total DNA of FUS ECR constructs were maintained to the recommended range that adopted from (Hogan et al., 1986). Reporter gene constructs were co-injected with a plasmid containing chick β-actin promoter driving td Tomato expression as an internal control for successful electroporation. In order to allow the reporter gene expression to develop fully, and to

allow differentiation of motor neurone cells, the expression was not assessed until 48 hours post transfection. The 25 successfully transfected embryos number from the three experiments had strong td Tomato in the neural tube (Figure 3.16A) and strong GFP in the same location (Figure 3.16B). Frozen sections of one of the transfected embryos showed the extent of successful electroporation in the left-hand neural tube since the DNA within this neural tube is directed to the cathode (Figure 3.16D). Figure 3.16E shows that the ECR PP can drive GFP expression in all the electroporated cells, including the developing motor neurones. These frozen sections were stained with DAPI (Figure 3.16F). Some frozen sections of the transfected embryos were also immunohistochemically stained using an Hb9 antibody (obtained from the University of Iowa) and a secondary antibody of goat anti mouse IgG Alexa 350. The stained section demonstrates Hb9 expression; specifically in the nuclei of motor neurone cells in the frozen section of the E5 chick embryo neural tube the Hb9 antibody stain shows a pale blue approaching violet fluorescence. Since the staining here was quite pale, however (Figure 3.16I) and since it was even harder to see with red and green fluorescence (Figure 3.16G and 3.16H), this experiment was not repeated for specimens from tissues that were transfected with other FUS ECRs constructs.

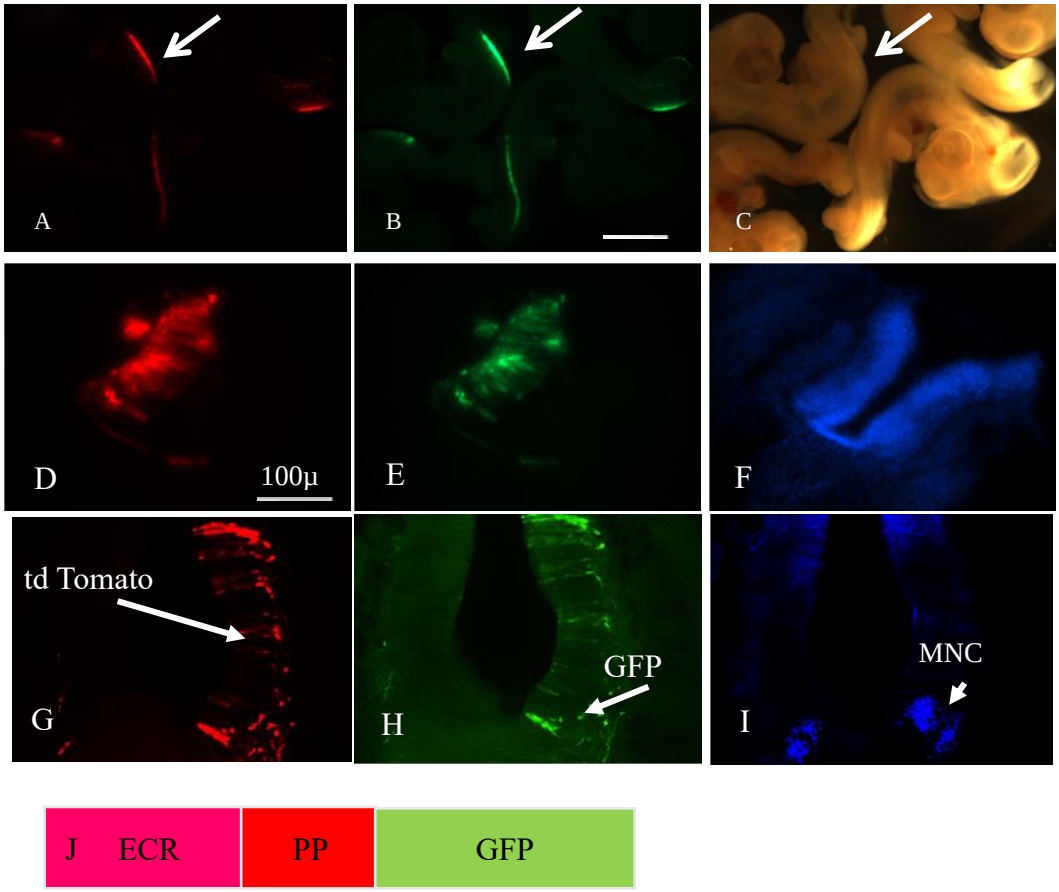


Figure 3.16: FUS ECR PP phrGFP activity in chick embryos. (A, B) and (C) show embryos that have been electroporated with 1 µg/µl td Tomato, 3.6 µg/µl ECR PP GFP, (A) td Tomato fluorescence indicating successful electroporation. (B) GFP fluorescence indicating the activity of ECR PP in the neural tube. (C) Bright field showing the location of electroporation in four successfully electroporated chick embryos. The arrow indicates the electroporated trunk region of one of the embryos. Scale bar 2 mm. D, E, and F are frozen sections of one of the embryos: D shows the extent of successful electroporation in the neural tube. It is restricted to the left hand side since the DNA within the neural tube is directed to the cathode. (E) Shows that ECR PP drives GFP expression in all the electroporated cells including the developing motor neurons. (F) Shows Dapi staining.(G, H, I) are the same frozen cross-section showing td Tomato (G), (H) motor neuron cells transfected with ECR PP phr GFP reporter gene, indicted by the white arrow (I) motor neuron cells demonstrating Hb9 expression specifically in the nuclei of motor neurons (J) Schematic representation of FUS ECR PP phrGFP reporter gene vector. Scale bar 100 µm

3.3.16 FUS PP phr GFP cannot drive reporter gene expression in chick embryo neural tubes

The functional activity from FUS PP phr GFP plasmid was tested in chick embryo neural tubes (Figure 3.17F). 1 µl each of FUS PP phr GFP at (11 µg/µl), of tomato plasmid (1 µg/µl), and of Fast Green (1 µg/µl) were transfected into the lumen of chick embryo neural tubes at stage HH 12-14. The ten successfully transfected embryos had strong red expression in the neural tube (Figure 3.16A) but no GFP expression was observed in any of the transfected samples, indicating that PP alone could not drive the detectable reporter gene (Figure 3.17B). Frozen sections from one of the embryos that were transfected with the FUS PP shows td Tomato expression, indicating successful transfection (Figure 3.17D). Figure 3.17E, meanwhile, shows no GFP expression in the neural tube, indicating that FUS PP alone cannot drive reporter gene activity in chick embryo neural tubes. Figure 3.17F Schematic representation of the FUS PP phr GFP constructs.

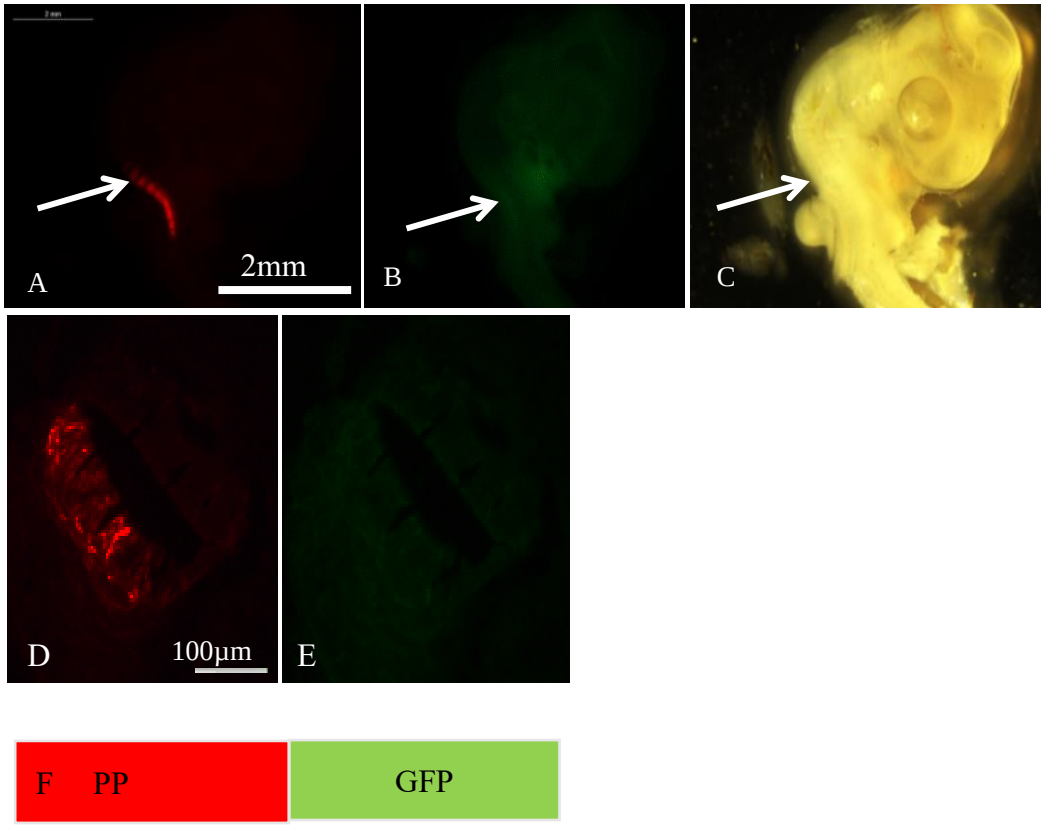


Figure 3.17: FUS PP phrGFP activity in chick embryo neural tube. All panels show embryos that have been electroporated with 1 µg/µl td Tomato, 3. 6 µg/µl PP GFP. (A) td Tomato, (B) PP GFP and (C) bright field microscopy image.. Scale bar 2 mm. The white arrow indicates the electroporated trunk region of the embryo. (D, E) are frozen sections of the embryo from (A, B and C). (D) Shows the extent of successful electroporation in the neural tube. It is restricted to the left hand side since the DNA within the neural tube is directed to the cathode. (E) Shows no fluorescence, indicating that PP alone cannot drive GFP expression in the neural tube, including in the developing motor neurons. (F) Schematic representation of FUS PP phrGFP reporter gene vector. Scale bar 100 µm.

3.3.17 FUS INT PP phrGFP acts as a positive activator in chick embryo neural tubes

The FUS INT PP 1phrGFP plasmid shown in (Figure 3.18F) was tested in chick embryo neural tubes. 1 µl each of FUS INT PP phr GFP at (3.6 µg), of tomato plasmid (1 µg/µl), and of Fast Green (1 µg/µl) transfected to the lumen of chick embryo neural tubes. After 48 hours the successfully transfected embryos were showed to have td Tomato expression when observed under the red filter and showed GFP expression under the GFP filters. In the three experiments 14 embryos had strong td Tomato (Figure 3.18A) and strong GFP expression in the chick embryo neural tubes at the same location (Figure 3.18B). Figure 3.18 D, E, meanwhile, show sections of the embryo neural tube transfected with the INT PP phrGFP. These section have been cut with a lab blade to observe the expression activity in more detail, Figure 3.18D shows td Tomato expression, indicated by the white arrow, and Figure 3.18E shows expression of IN PP phrGFP, indicated by the blue arrow. Motor neurone cell locations are indicated by a purple arrow.

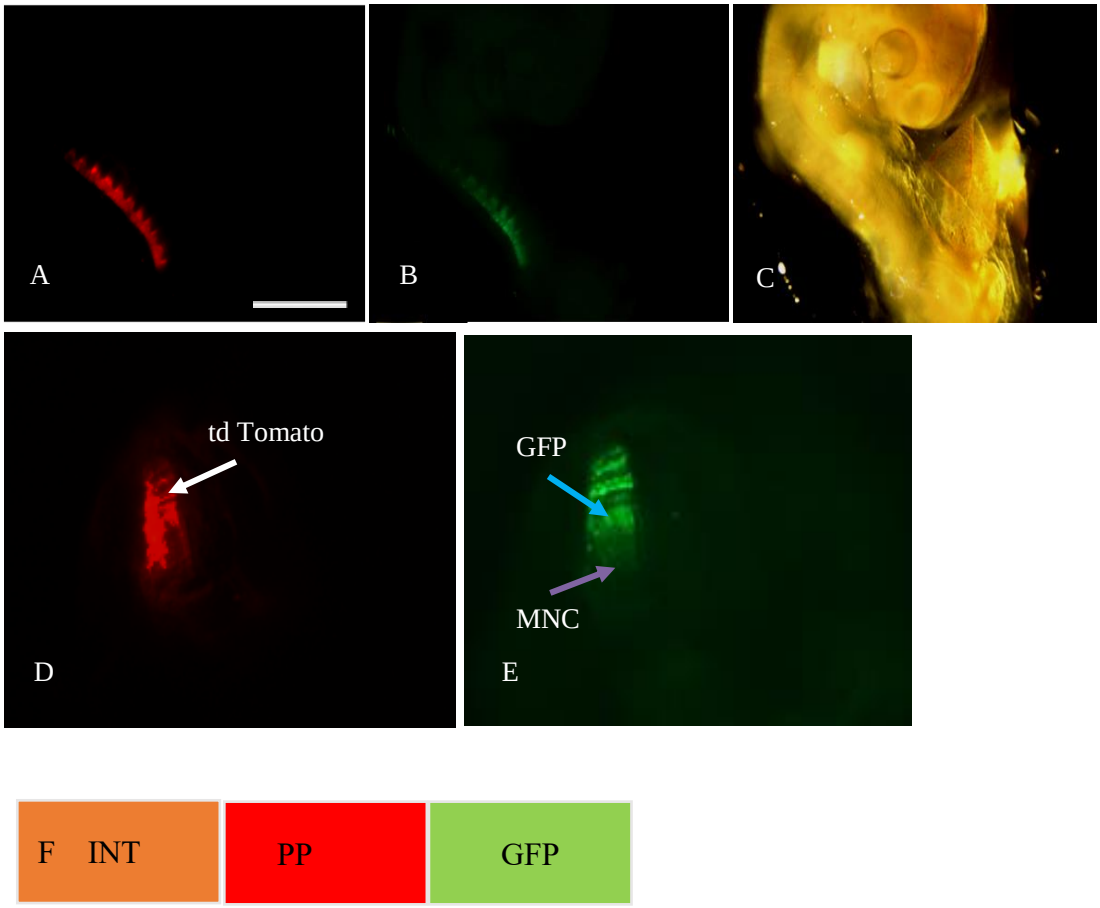


Figure 3.18: FUS INT PP enhances expression in chick embryo neural tube. *(*A, B and C) show embryos that have been electroporated with 1 µg/µl td Tomato, 3.6 µg INT PP GFP. (A) td Tomato fluorescence indicating successful electroporation. (B) GFP fluorescence indicating the activity of INT PP in the neural tube. (C) Bright field microscopy image. (D, E) Cross-sections of the neural tube from the transfected embryo in (A), (D) shows td Tomato in the neural tube cells and (E) shows FUS INT PP activity indicated by GFP in most neural tube cells, indicated by the white arrow and including motor neuron cells indicated by the purple arrow (F) Schematic representation of FUS INT PP phrGFP reporter gene vector. Scale bar 1 mm

3.3.18 FUS INT ECR PP phrGFP can act as activator in chick embryo neural tubes

FUS INT ECR PP phr GFP was tested in chick embryo neural tubes (Figure 3.19F). The test plasmid mixture at a final concentration of 4.4 µg, 1 µg/µl of td Tomato, 1 µg/µl Fast Green, and 1 µl of PBS was injected and transfected to the chick embryo neural tubes. 48 hours post transfection the 14 successfully transfected embryos showed both strong td Tomato (Figure 3.18A) and strong GFP expression (Figure 3.19B) in the same location of the neural tube. Frozen sections were taken from the neural tubes of transfected embryos and Figure 3.19D shows the expression of td Tomato in these sections, indicated by white arrow, while Figure 3.19E shows the expression of the INT ECR PP in the transfected area, indicated by the blue arrow; motor neurone cells are indicated by the purple arrow.

Overall, the promoter/ enhancer properties of these various regulatory regions (ECR PP, INT PP, and INT ECR PP) were consistent with the results obtained in the cell line model, except that PP alone showed no activity in vivo. Differences of FUS PP behave between SK-N-AS and chick embryo neural tube at E5 is most likely due to transcription factors that are active *in vitro* but not active *in vivo* at this development stage of neural tube ; or they might have a tissue specific activity and not active in the developing neural tube cells. These differences could play a crucial role in the response to regulatory elements.

The transfected neural tubes were excised and fixed and frozen cross-sections were prepared. In the FUS ECR PP phr GFP transfected specimen, the majority of GFP positive cells were located in the ventral region, specifically in the motor column and motor neurone cells. In the INT PP and INT ECR PP specimens, strong GFP signals

were apparent in the dorsal and ventral regions including the motor column. PP alone did not show fluorescence in the dorsal or ventral axes.

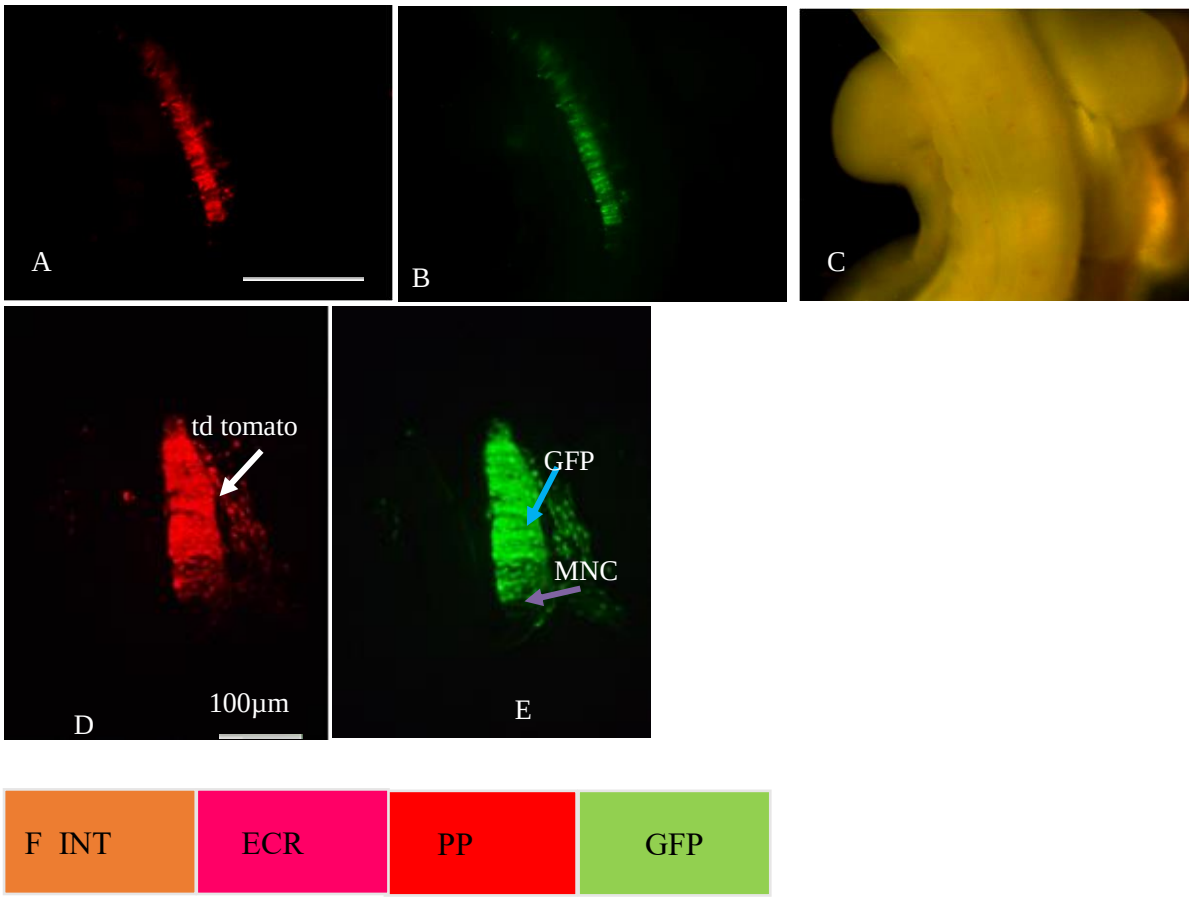


Figure 3.19: FUS INT ECR PP activity in chick embryos. (A, B, and C) show embryos that have been electroporated with 1 µg/µl td Tomato and 4.4 µg/µl INT ECR PP. (A) td Tomato fluorescence. (B) GFP fluorescence indicating the activity of INT ECR PP in the neural tube. (C) Bright field microscopy image. Scale bar 2 mm (D) frozen section of transfected neural tube showing td Tomato expression (E) the same section showing expression of GFP in the neural tube and specifically in the motor neurone cells MNC (F) Schematic representation of FUS INT ECR PP phr GFP reporter gene vector. Scale bar 100 µm.

3.3.19 Transfection of S K- NA-S cell line with FUS regulatory domains

The activities of the predicted regulatory regions from the FUS gene were assessed in the FUS-expressing neuroblastoma SK-NA-S cell line. This experiment was conducted particularly to test whether FUS PP plasmid alone was constructed correctly and to affirm that there was no problem with the construction of the plasmid. The activity of the constructs was measured using the hr GFP reporter gene system as a traceable marker and a scoring system based on the number of cells expressing GFP. The td Tomato was included as a positive control and promoter-less phrGFP alone as negative control. Construct activity was determined 24 hours after transfection by counting the number of GFP-positive cells as a proportion of the total number of cells, as viewed by fluorescent and phase-contrast microscopy, respectively. All predicted regulatory regions showed significant marker gene expression (Figure 3.20), table 5 in the appendix. The cells transfected with INT ECR PP phrGFP showed the highest level of GFP expression among all FUS ECR constructs, while cells transfected with ECR PP phrGFP showed the lowest level of GFP expression. Cells transfected with INT PP phrGFP and PP phrGFP constructs expressed similar level of GFP expression.

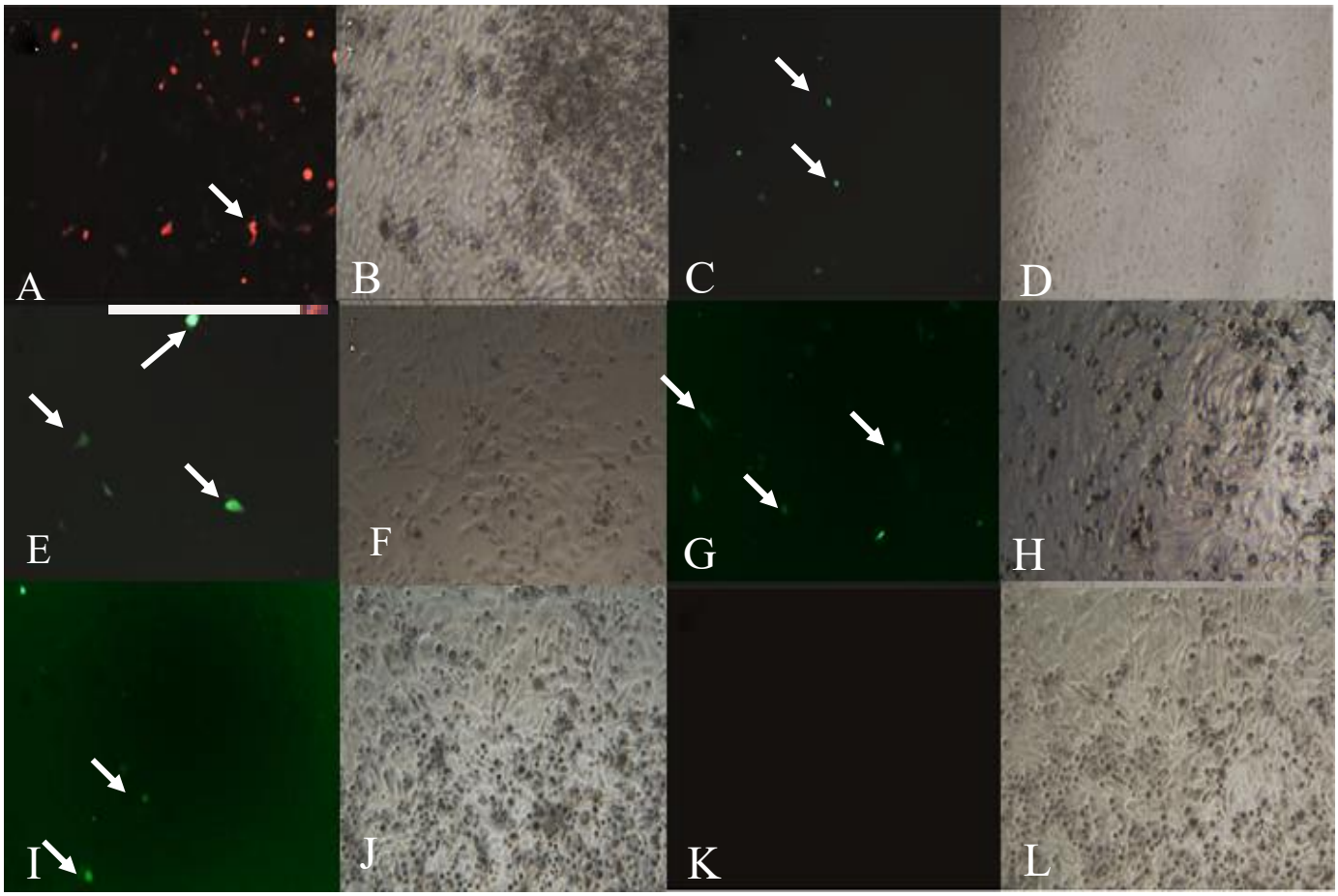


Figure 3.20: Transfection of SK-NA-S cells with FUS ECR phrGFP constructs. The red cells in the fluorescence image (A) are transfected cells with td Tomato plasmid represented as a positive control (B) phase picture of the same image. The green cells in the fluorescence (C) are transfected cells with FUS ECR PP construct (D), phase picture of the same image. The green cells in the fluorescence (E) are transfected cells with FUS PP (F) phase picture of the same image. (G) Fluorescence of cells transfected with FUS INT ECR PP (H) phase pictures of the same image. (I) fluorescence of cells transfected with FUS INT PP (J) phase pictures of the same image. (K) No fluorescence is seen in cells transfected with phr GFP alone, represented as a negative control (L) phase pictures of the same image. Scale bar 500µm

3.4 Discussion

The LacZ reporter system works well in the chick embryo model but the long processing time meant that it was not an ideal technique for the current study. It was therefore replaced with an alternative reporter gene system using GFP and td Tomato markers. The GFP plasmid functioned as a reporter gene, whereas td Tomato functioned as an internal electroporation efficiency control. Furthermore, electroporation drives the DNA in one direction and GFP was not expressed at the site contralateral to injection, demonstrating successful microinjection and electroporation. This dual fluorescent protein system has many advantages including: the transfected embryos can be observed under a fluorescent microscope without embryo sacrifice; the GFP fluorescence spectrum did not overflow into red filters and vice versa; embryos can be cultured *in ovo* for 48-72 hrs or much longer therefore gene expression can be monitored over a longer time scale with no leakage of proteins into surrounding tissues or background; and, finally, gene expression is directly visualised under fluorescent microscopy without the need for staining.

Comparative genome analyses helped in the identification of domains in the human FUS gene that have been conserved across species which we postulated were potentially important regulatory domains. This analysis identified putative domains including a 704 bp domain located in the first intron of the FUS gene, a minimal promoter domain contiguous with MTSS, and a 311 bp ECR fragment upstream of the proximal promoter (see Figure 3.11).

These regulatory regions were tested with SV40 minimal promoter in the neuroblastoma cell line, SK-N-AS. ECR and INT were shown to act as transcriptional

activators and drove expression of the luciferase marker gene in this cell line for example ECR increased the level of expression activity by two fold and INT1 increased the level of expression four time more (Figure 3.12B). Same regions were tested with hybrid neuroblastoma/motor neuron cell line NSC-34 cells, ECR increased the level of activity two fold and INT1 showed to increase the level of activity 8 time more (Figure 3.12C).

Both INT and ECR when linked with FUS PP in the phrGFP plasmid had transcriptional activity in the chick embryo neural tube, INT PP expression were observed in 14 transfected embryos and ECR PP expression were observed in 25 transfected embryos, while PP alone had no activity in the *in vivo* environment. When the same plasmids were tested *in vitro* they show GFP expression, and PP alone shows enhancer activity. These experiments were conducted to ensure that the plasmids were constructed correctly; they showed that there is no difference in the plasmid construction but that there are differences between the chick and SK-N-AS environment and / or sensitivity of detection. The difference is most likely due to differences in the cellular environment or phenotype; for example SK-N-AS cells are human cancer, transformed cells with a lot of mutations and might have different transcription factors. The differences might also be due to transcription factors that are active *in vitro* but not active *in vivo*; or they might have a tissue specific activity and not transcript in the developing neural tube cells. These differences could play a crucial role in the response to regulatory elements.

Identification of the regulatory domains that regulate FUS gene expression in motor neurones can help us to understand the genetic features that may promote these diseases. For example, the INT domain was shown to have enhancer activity in the *in vitro* experiment with the pGL3b vector (Figure 3.12A). The INT domain in combination with PP, when maintained after MTSS, was shown to affect the transcriptional activity of the luciferase marker gene and act as an activator of transcription; specifically, the INT domain when located after MTSS can modulate through post-transcriptional activity and demonstrate a significant increase in the transcription level. When it has been cloned upstream of PP (Figure 3.21) or upstream of ECR PP (Figure 3.22) meanwhile, strong GFP expression can be detected in the chick embryo neural tube serving as a transcriptional activator.

Sequences 1

ATGGCCTCAAACGGTAGGTAAGGGCGCGAGGCGACGGCGGCGGCGCACCCGGCCGAGG
CCTCCCAGCTGGGCTTTTCGTTTTCAGTGGGACCGGGGCGGCGATCCCGTGTGGGATTTT
TTGGCGCCCCTGTGGCGGGAAGCCGCGGAGAAGAGTAACTGGAGGAGGCTGGTGTCGC
CATTTTGTTTCGCTCCTCTGGCCCTCGCGCGCGGGGCGGGAAGTCTTTTCTTTGCAGTCC
GTTTGCTTGGGGTGGGCGTTGGGAGGGACGCTTCTTAGGGGTTTGAAGCGTCAGGTGAG
GGTGGAAAACGCCCATTCTCCGTGGCCTCGCCTCCCCCAACTCCCGGCCCCGCGCTCGA
GCCCGCTTTGTCGCAGTGCTGCATCCGGGCACTCGCGGCGCGCACGCGCTCTGCGGGCC
CCTCCCCCTTCGCGGCGCGGGTACCCCTTCCCCGCCTCGTGTTGGTTCAGCTTTCTGTCG
CGAGACCCTTCGCGGAAGACTCGGCGGCGCGCGTCCGGTGTGAGCCTTGTCCCTCAGTG
GTCCTTCGCGAATGGGCGGGACCGCTCCGTTCCCGCCTGGGTTGCCACGCGGCTGGGGG
CGGAGGCTCGGGATCGGGGCCGCCCTCTAGCTTAACGGTTTGGCGGCGGTGGTCAGGGT

Sequences 2

TCGACCAACGGACTTGGGGACGGCCCGAGAGTTTTTCCCGCCTAAATTTCTTTCGACGAGGG
AGGCTGGAGCCGCGGGGACGAGGCGCCCCATACAGCGGCAAGAGGGTGGAGGGCAGGA
GCTCGCCATCCTGGGTGAAAGCGGGGCCCAGCGAAGGGGCCCGGCCACAGGAATCTCGG
TTCCACCCCGCTACTCCCGGCTGTGACTCCAGTTTCGTCCCCAGCCGCCGGGACCGCCCCC
TCGCCCCGCCCCCAGCGGGCACTCAGGCCGTACCACTGTGCCTTCATGGGGGTGGAGATA
GATCGTGGGCTAGTCCTGCCGAGGAGAGAGGGGTTCTTCCTCAAAAAATATGATTATGTA
TAGTATTCGCATGATTCTAGTTAACTTGTTTCCCTTCTGCCTGCTCGGACCCTCTACCTGCC
CTACGAAGGGGGCGGAGTGCGTTCCTGCCTCCCCCTGCTCTTCCGCGTTTGGTGCGCGCCT
GCGCGGTGCGTAGGCGGCGGAGCGTACTTAAGCTTCGACGCAGGAGGCGGGGCTGCTCA
GTCCTCCAGGCGTCGGTACTCAGCGGTGTTGGAACTTCGTTGCTTGCTTGCCT

Figure 3.21 FUS INT PP sequences. INT1 sequences from human FUS gene labelled with sequences1, were cloned upstream of FUS promoter labelled with sequences 2, red font of sequences1 indicates the start of INT1sequences, red font of sequences2 indicates the start of FUS promoter.

Sequences 1

ATGGCCTCAAACGGTAGGTAAGGGCGCGAGGCGACGGCGGCGGCGCACCCGGCCGAGGCCTCCCAGCTGGG
CTTTTCGTTTTCAGTGGGACCGGGGCGGCGATCCCGTGTGGGATTTTTTGGCGCCCCTGTGGCGGGAAGCCGC
GGAGAAGAGTAACTGGAGGAGGCTGGTGTCGCCATTTTGTTTCGCTCCTCTGGCCCTCGCGCGCGGGGCGGG
AAGTCTTTTCTTTGCAGTCCGTTTGCTTGGGGTGGGCGTTGGGAGGGACGCTTCTTAGGGGTTTGAAGCGTCA
GGTGAGGGTGGAAAACGCCCATTCTCCGTGGCCTCGCCTCCCCCAACTCCCGGCCCCGCGCTCGAGCCCGCT
TTGTCGCAGTGCTGCATCCGGGCACTCGCGGCGCGCACGCGCTCTGCGGGCCCCTCCCCCTTCGCGGCGCGG
GTACCCCTTCCCCGCCTCGTGTTGGTTCAGCTTTCTGTCGCGAGACCCTTCGCGGAAGACTCGGCGGCGCGCG
TCCGGTGTGAGCCTTGTCCCTCAGTGGTCCTTCGCGAATGGGCGGGACCGCTCCGTTCCCGCCTGGGTTGCCA
CGCGGCTGGGGGCGGAGGCTCGGGATCGGGGCCGCCCTCTAGCTTAACGGTTTGGCGGCGGTGGTCAGGGTT
CGACCAACGGACTTGGGGACGGCCCGAGAGTTTTTCCCGCCTAAATTTCTTTCTAGAGCTCGCCATCCTGGGT
GAAAGCGGGGCCCAGCGAAGGGGCCCGGCCACAGGAATCTCGGTTCCACCCCGCTACTCCCGGCTGTGACT
CCAGTTTCGTCCCCAGCCGCCGGGACCGCCCCCTCGCCCCGCCCCCAGCGGGCACTCNNCCGTACCACTGTG
CCTTCATGGGGGTGGAGATAGATCGTGGGCTAGTCCTGCCGAGGAGAGAGGGGTTCTTCCTCAAAAAATATG
ATTATGTATAGTATTCGCATGATTCTAGTTAACTTGTTTCCCTTCTGCCTGCTCGGACCCTCTACCTGCCCTAC
GAAGGGGGCGGAGTGCGTTCCTGCCTCCCCCTGCTCTTCGCGTTTGGTGCGCGCCTGCGCGGTGCGTAGGCG
GCGGAGCGTAGCGTACTTAAGCTTCGACGCAGGAGGCGGGGCTGCTCAGTCCTCCAGGCGTCGGTACTCTAG
CGGTGTTGGAACTTCGTTGCTTGCTTGCCTGTGCGCGC

Sequences 2

Sequences 3

Figure 3.22 FUS INT1 sequences cloned upstream of FUS ECR PP in reporter plasmid phrGFP. INT1 labelled with sequences1, were cloned upstream of FUS ECR promoter labelled with sequences 2 and sequences 3 for FUS promoter, red font of sequences1 indicates the start of INT1sequences, red font of sequences2 indicates the start of FUS ECR, red font of sequences 3 indicates the start of FUS promoter.

It has been shown that mutation within cis regulatory domains or 3'UTR mutations of FUS gene may have a profound effect on the gene expression spatially and temporally that might lead to changes in FUS gene location in the gene in the cell and cause globular and granular inclusions, eventually leading to loss of motor neurones (Sabatelli et al., 2013a, Gal et al., 2011, Kryndushkin et al., 2011, Ito et al., 2011). These ECR domains may now be added to the list of regions of the gene to be considered for study when undertaking mutational analysis of samples from patients with MND. Furthermore, they are also regions of interest in view of recent results that correlated regulation of FUS expression with motor neurone degeneration (Vance et al., 2013).

The results presented here suggest that chick embryos can be used as a model system to confirm the role of the human FUS regulatory domains in the motor neurones of chick embryo neural tube. These human regulatory domains were selected on the basis of their conservation across phyla from fish to chick to man. The approach used here can be applied to analysis of other genes whose expression is implicated in the development and progression of ALS and other neurodegenerative diseases (Abe et al., 2004)

Although the FUS ECRs used are human sequences and their conservation in chicks is not known, they have responded well to the chick environment in vivo and, specifically, their activities were observed in the motor neurones of the neural tube. These results demonstrate that the chick embryo model can be used to investigate potential regulatory elements that may function in the adult human diseases.

In summary, this chapter has demonstrated a novel human regulatory domain, INT, located within the MTSS of the FUS gene, which could play a role in FUS gene regulation in the form of ECR. The positive results generated from the FUS domain's activity in the chick embryo neural tube and the success achieved from using this model system provides an incentive to test other putative regulatory domains that are located upstream of two genes FUS and PARK7. These domains are SVAs and are classified as subtype D, a subtype that is only found in primates and is not conserved in chicken Savage et al. (2014b).

Chapter 4

The FUS SVA and PARK7 SVA can act as transcriptional regulators

4.1 Introduction

In the previous chapter it was demonstrated using a chick embryo model that the regulatory domain resides in the evolutionary conserved regions (ECRs) of the human FUS gene. In this chapter, the study is extended to investigate the role of another domain or region located upstream of the FUS gene. This region is a SVA, and has been demonstrated to be important in FUS gene regulation (Savage et al., 2014a). This region was identified using the genome browser UCSC and ENCODE data (the Encyclopaedia of DNA elements). This aided the identification of the regulatory elements that might have an influence on FUS gene regulation. Specifically, when the UCSC genomic browser (http://genome.ucsc.edu/index.html) was applied to the FUS gene locus, a large repetitive region was identified located within 10 kb of the transcription start site (TSS) from the FUS gene and 20 kb away from the 5' end of the PRSS36 gene. This repetitive region was called VNTR and was found to be a central region of a large repetitive region known as SVA (see Figure 4.1). The ENCODE data further identified many elements that might suggest that this SVA could be involved in gene regulation. These are: 1) DNase 1 clusters were found on both sides of the SVA. 2) regions with active histones and H3K4Met were found, which are always found near to the transcription factor binding and 3) human expressed sequences tags (ESTs) were found to originate and be transcribed in both directions of this region (Figure 4.3) (Doolittle, 2013, Kavanagh et al., 2013). SVAs are non-autonomous, non-LTR retrotransposons, named according to their composition SINE-R, VNTR (variable number of tandem repeats), and an *Alu*-like sequence (Wang et al., 2005). SVAs are actively mobilised in the human genome, using the LINE 1 machinery in order to replicate and expand through the genome by a copy and paste mechanism (Wang et al., 2005). 2700 copies of SVAs are present in

the human genome (Wang et al., 2005). The general structure of the SVA and the classification of its subtypes were described in the Introduction chapter, section 1.8.3. The FUS SVA is classified as subclass D which is found only in humans and chimpanzees. The structure of FUS SVA differs to the structure of the canonical SVA by the absence of two elements (see Figure 4.1). First, FUS SVA has no hexamer repeat CCCTCT at the 5' end. Second, it has no poly A-tail after the SINE. FUS SVA has an Alu–like sequence, and a central repetitive region consisting of tandem repeat (TR) and VNTR and followed by SINE. The structure of the canonical SVA is illustrated in Figure 1.8 in the Introduction chapter

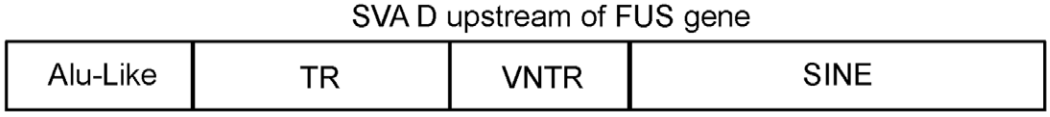


Figure 4.1: Structure of the SVA located upstream of the FUS gene.
A diagram illustrating the structure of the SVA D located upstream of FUS gene. It consists of an Alu-like sequence, then a TR followed by VNTR and a SINE. FUS SVA is missing the hexamere repeat CCCTCT at the 5' end of a canonical SVA and has no poly A-tail after the SINE, this image is taken from (Savage et al., 2014b).

The estimated rate of the insertion of SVAs insertion into the human genome is 1 in every 916 births (Xing et al., 2009)). To date, it is known that eight of the inserted SVAs have a correlation with diseases (Hancks and Kazazian, 2012). Global analysis of the SVAs has identified another SVA located upstream of the PARK7 gene (Figure 4.2) (Savage et al., 2013a). This SVA is important for, the fact that it is located upstream of the PARK7 gene is significant since the gene is known for its association to neurodegenerative disease and in some cases ALS (Buée-Scherrer et al., 1995, Hof et al., 1994) and also in two types of breast cancer, and in non-small cell lung carcinoma (MacKeigan et al., 2003a). PARK7 SVA shares similarities with the FUS SVA feature as it is located 10 kb upstream of PARK7 gene, it is human specific and

is classified as sub class D. PARK7 SVA, however, differs from FUS SVA in that it contains all the domains of the canonical SVA, for example it has a CCCTCT hexamere repeat at 5' end.

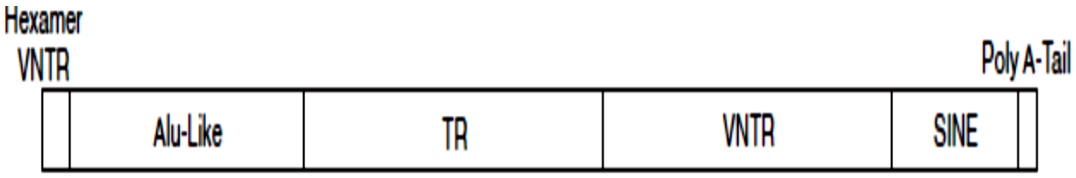


Figure 4.2 the human specific PARK7 SVA located 10 kb upstream, of the PARK7 gene. SVA has a hexamer repeat CCCTCT, VNTR, Alu-like sequence, TR, VNTR, SINE and poly A-tail (redrawn from (Savage et al., 2013a)

In summary, previous work has identified two SVAs, located upstream of the FUS and PARK7 genes as regulatory elements in an *in vitro* model (Savage et al., 2014b), (Savage et al., 2013a). This chapter will demonstrate the potential functional activity of both FUS SVA and PARK7 SVA *in vivo* using the chick model established and validated in chapter 3 to determine if these elements have regulatory functions.

4.2 Aims

- Analyse the endogenous expression of chick PARK7 gene in chick embryo brain at E5.
- To study in the developing chick central nervous system the regulative properties of the SVAs in FUS and PARK7 genes known for their association with neurodegenerative disease. This will be achieved by cloning the sequences of the SVA and VNTR from FUS gene, SVA from PARK7 and insert into a reporter plasmid; then to inject the generated plasmids into neural tube, midbrain region of developing chick embryo at E2-3 and then electroporated to test their regulation activity by driving the expression of phr GFP

4.3 Methods

- Bioinformatics analysis of the FUS, PARK7 gene locus in the UCSC genomic browser (http://genome.ucsc.edu/index.html) was used in order to provide information about the function and structure of important repetitive regions that have potential regulatory function.
- Develop a methodology in order to inject and electroporate chick embryo mid brain at E5 with FUS SVA, PARK7 SVA constructs and analyse their regulatory activities using reporter gene system phrGFP

4.3.1 Construction of human FUS SVA and VNTR reporter plasmid domains for *in vivo* experiments

The generation of FUS proximal promoter reporter plasmid has been described in Chapter 2, section 2.15.15.1. The FUS SVA and FUS VNTR plasmids were provided by Dr Vivien J. Bubb University of Liverpool). They contain the intact FUS SVA and the isolated TR/VNTR sequences and were PCR amplified from SVA and VNTR reporter plasmids used for the *in vitro* analysis and cloned upstream of the promoter into either FUS PP phr GFP or FUS ECR PP phr GFP, which have already been used in Chapter 3 section 3.3.15 and section 3.3.16 The resulting constructs were named SVA PP phrGFP and VNTR PP phrGFP SVA ECR PP phrGFP, VNTR ECR PP phrGFP. The constructions of these four plasmids were described in Chapter 2, section 2.15.15.3

4.4 Results

4.4.1 Identification of a SVA D upstream of FUS gene

When the UCSC genome browser analysis was applied on the FUS gene locus aided by utilisation of ENCODE (the Encyclopaedia of DNA elements) it illustrated the structure of the gene in the region under examination, highlighting an active chromatin area and binding site of transcription factors (Figure 4.3). A large GC- rich repetitive region located 10 kb away from FUS gene transcriptional start site (TSS) was identified when the FUS gene loci were analysed. The ENCODE data suggested that this repetitive region had many features indicating that it might be active; these features included: 1) being flanked by a DNase 1 hypersensitivity cluster, and 2) the transcription factor binding sites. This repetitive region was found to be part of larger repetitive region called an SVA (Figure 4.3). This SVA is present in humans and chimpanzees only and was classified as a SVA subtype and conservations were observed between species (Figure 4.3).

Figure 4.3: UCSC genome browser image illustrating upstream of the FUS gene. In the UCSC genomic browser the scale and chromosomal loci are located at the top of the image and the tracks names are located on the left hand side of the display. The illustrated region can be chosen by the user; in this image from the top it can be seen that the transcription factor binding is overlapping the location of the repetitive region. SVAD and VNTR region sequences are boxed in blue.* An evolutionary conserved region which are not conserved in chickens, H3K27Ac histone mark (often found near regulatory elements), DNase clusters, transcription factor binding.

4.4.2 FUS SVA D, VNTR act as transcriptional regulator in reporter gene constructs *in vitro*

The functional analysis of the FUS SVA or VNTR *in vitro* was addressed in the laboratory by Dr Thomas Wilm. The data generated by the experiments have been summarised here to demonstrate the functional properties of both FUS SVA and VNTR *in vitro* (Khursheed et al., 2015). The regions encompass the SVA, and within the central VNTR were PCR amplified and cloned in the firefly luciferase reporter gene pGL3P, a vector containing a SV40 minimal promoter element. These constructs containing the complete SVA sequences or the VNTR were tested in the human neuroblastoma cell line SK-N-AS. The level of reporter gene expression showed significant differences between SVA and VNTR compared to the minimal promoter SV40 promoter alone in the pGL3P vector. The complete SVA repressed reporter gene and showed decreased in the level of expression by $p<0.05$, whereas the TR/VNTR were showed enhanced properties and increased the level of reporter gene expression by $p<0.001$ for short alleles and by $p<0.05$ for long alleles in this cell line. It may demonstrate that SVA may contain distinct regions that act as regulatory domains and one of these domains act as a repressor in SK-N-AS cells.

4.4.3 FUS SVA and VNTR can act as regulatory domains in the chick embryo neural tube

Previous chapter we have discussed the activity of the FUS ECR in chick embryo neural tube and showed these regions have a regulator activity. Here, the functional activity of SVA and VNTR from FUS gene will be tested in chick embryo neural tube. Interestingly, previous work by (MacKenzie and Quinn, 1999) showed that specific VNTRs can have a regulator property and support tissue specific expression in developing mouse transgenic models. To investigate the repression activity of the FUS SVA in chick neural tube at development stage E5, the FUS SVA was analysed and cloned in to the FUS ECR PP plasmid, the expression properties of which were studied in detail in Chapter 3, and showed to act as a strong regulator. The ECRPP FUS SVA plasmid at 2-5 µg DNA at final concentration with 1 µg/µl of tomato plasmid were injected into the chick neural tube 48 hours after electroporation, strong tomato expression was observed in 6 embryos, (Figure 4.4A) table 4.4.15 specifically in the neural tube. When these samples were analysed for GFP expression, all were seen to also express GFP (Figure 4.4C).

Figure 4.4: Activity of ECR PP FUS SVA in chick embryo neural tubes at E5. (A) Embryos electroporated with tomato plasmid as marker for successful transfection. (B) Bright field microscopy image. (C) Expression of ECR PP FUS SVA in chick embryos' neural tubes indicated by reporter gene GFP (D) Schematic representation of ECR PP FUS SVA phrGFP reporter gene vector. Scale bar 1 mm

Next, tested plasmid was FUS SVA cloned into FUS PP plasmid. Previously FUS PP plasmid was observed, and shown to have no activity in chick embryos. In each transfection the total DNA of FUS SVA PP, FUS SVA ECR PP, FUS VNTR PP, and FUS VNTR ECR PP constructs were maintained to the recommended range that adopted from (Hogan et al., 1986). FUS SVA PP plasmid at (2-5)µg DNA concentration with 1 µg/µl of td tomato plasmid were injected into the chick neural tube, after 48 hours electroporation, FUS SVA PP expression was indicated by phrGFP and the electroporation efficiency was indicated by tomato reporter expression. td Tomato plasmid was found to be expressed strongly along the dorsal and ventral axis of the neural tube in 14 embryos (Figure 4. 5A)

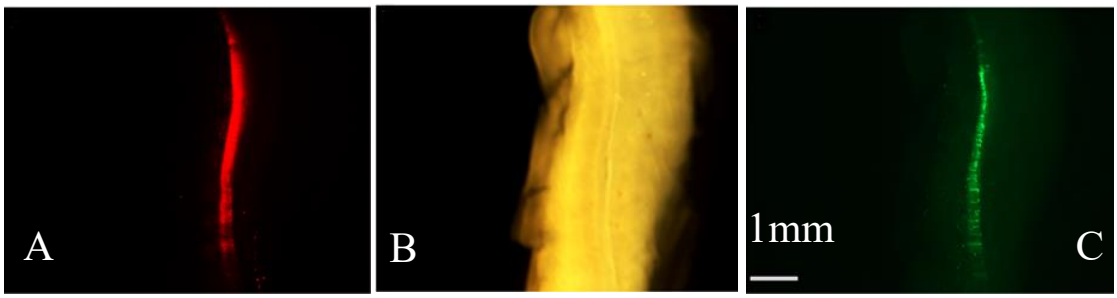


Figure 4.5: Representative image of FUS PP SVA phrGFP activity in the chick embryo neural tube at E5. (A) Expression of tomato plasmid in the chick embryo neural tube, indicating successful electroporation. (B) Bright field microscopy image. (C) Expression of the GFP reporter gene indicating activity of the FUS SVA as a transcriptional regulator in the chick embryo neural tube.

When analysed for GFP reporter gene expressions. Only four out of these 14 embryos were found to express the GFP reporter gene (Figure 4.6C), table 4.4.15. Figure 4.6C shows weak GFP expression in the neural tube of four embryos indicated by white arrows, it seems that SVA works in chick embryos as a weak regulator. Embryos show very strong td Tomato expression under the β-actin promoter (as in Figure 4.6A).

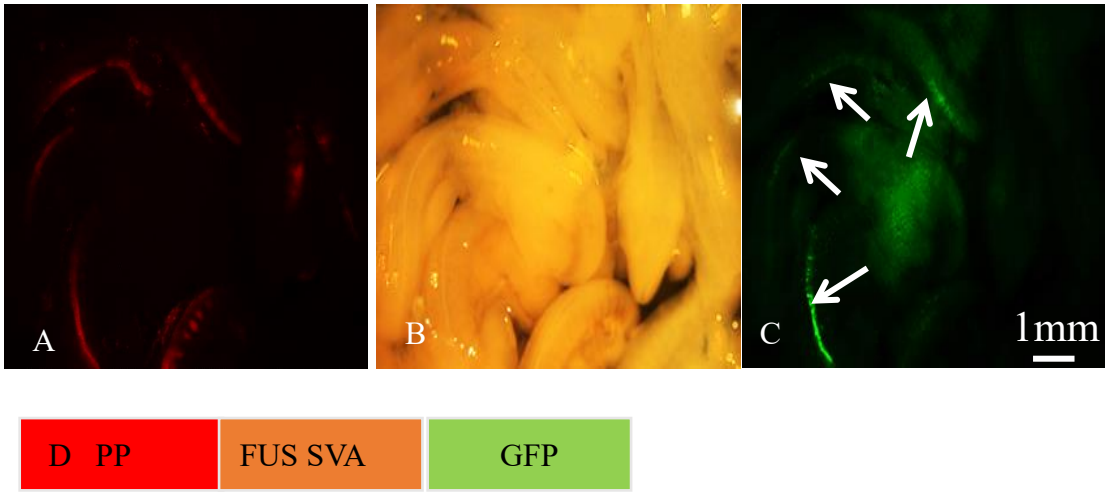


Figure 4.6: Representative image of the FUS PP SVA activity in chick embryo neural tubes at E5. Strong td tomato expression was observed in the chick neural tubes (A). (C) Weak GFP expression indicated by white arrows, showing that SVA can act as weak enhancer. (B) Same image in bright field microscopy. (D) Schematic representation of FUS PP SVA phrGFP reporter gene vector. Scale bar 1 mm

To test the activity of the FUS VNTR alone in the chick embryo neural tube, the same methods were followed as before and after 48 hours' incubation embryos were analysed for reporter gene expression. In contrast with the intact SVA, all embryos which expressed tomato plasmid strongly had GFP expression in the same location. All seven of the embryos that expressed tomato plasmid also expressed GFP in both the dorsal and ventral axis table 4.4.15.

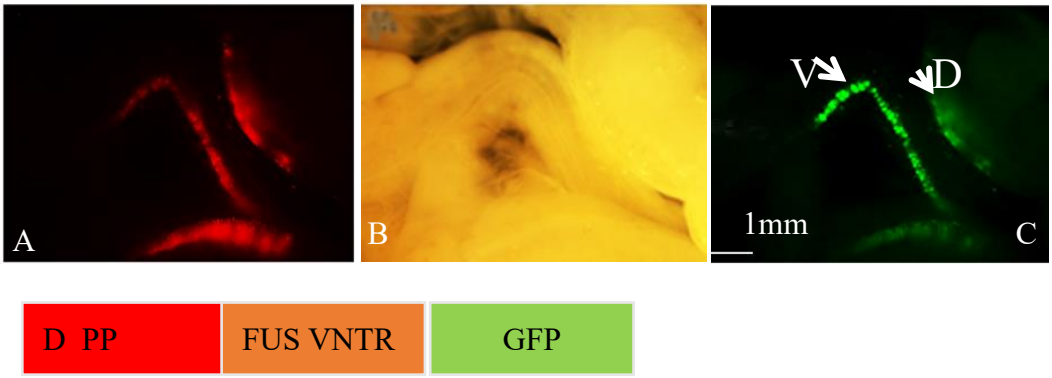


Figure 4.7: Representative image of the FUS VNTR activity in chick embryo neural tubes at E5. (A) Embryos electroporated with tomato plasmid as a marker for successful transfection. (B) Bright field microscopy image. (C) Expression of ECR PP FUS SVA in chick embryo neural tubes indicated by reporter gene GFP (D) Schematic representation of FUS VNTR phrGFP reporter gene vector. Scale bar 1 mm

.

FUS VNTR ECR PP phrGFP was tested to determine the VNTR transcription activity. Expression of the reporter gene construct was seen in the neural tube, and this expression indicated the strong enhancer activity of the FUS VNTR ECR PP phrGFP (Figure 4.8) by expression of GFP (C) and tomato plasmid (A) in the three electroporated embryos table 4.4.15. ECR PP is known for its strong activity (Khursheed et al., 2015) and this experiment confirmed that VNTR can elevate the expression activity and acts as an activator.

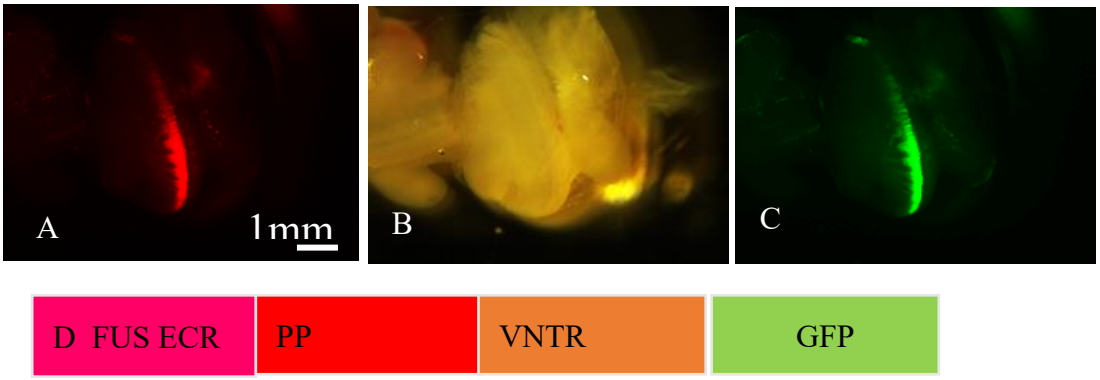


Figure 4.8: Representative image of FUS ECR PP VNTR expression in chick embryo neural tubes at E5. (A) Tomato plasmid (B) bright field microscopy image (C) Reporter gene expression FUS ECRPP VNTR (D) Schematic representation of ECR PP FUS VNTR phrGFP reporter gene vector. Scale bar 1 mm

.

4.4.4 Characterisation of a SVA located upstream of the PARK7 gene

The PARK7 SVA was identified during global analyses of SVAs in the human genome (Savage et al., 2013a). This SVA is located 10 kb upstream of the PARK7 gene. According to the repeat masker track on the UCSC genome browser (Hg 19) this SVA has features that are similar to FUS SVA as both are human specific, and subtype D, but differs from FUS SVA in having a hexamer repeat allocated at the 5' end. ENCODE data showed that active histone marks and DNase1 hypersensitivity clusters are located near to the promoter locus (Takahashi et al., 2001). Thus, this region has been identified as a major transcription start site. Another region was also identified 7 kb away from the TSS of the PARK7 gene. This region has similar features of identified TSS for PARK7 gene, including the CpG Island, DNase 1 hypersensitivity clusters and active histone marks, and this region is therefore referred to as a minor transcription start site. Furthermore, a human specific SVA was identified 8 kb upstream of the MTSS and 1 kb upstream of the minor TSS (Figure 4.9).

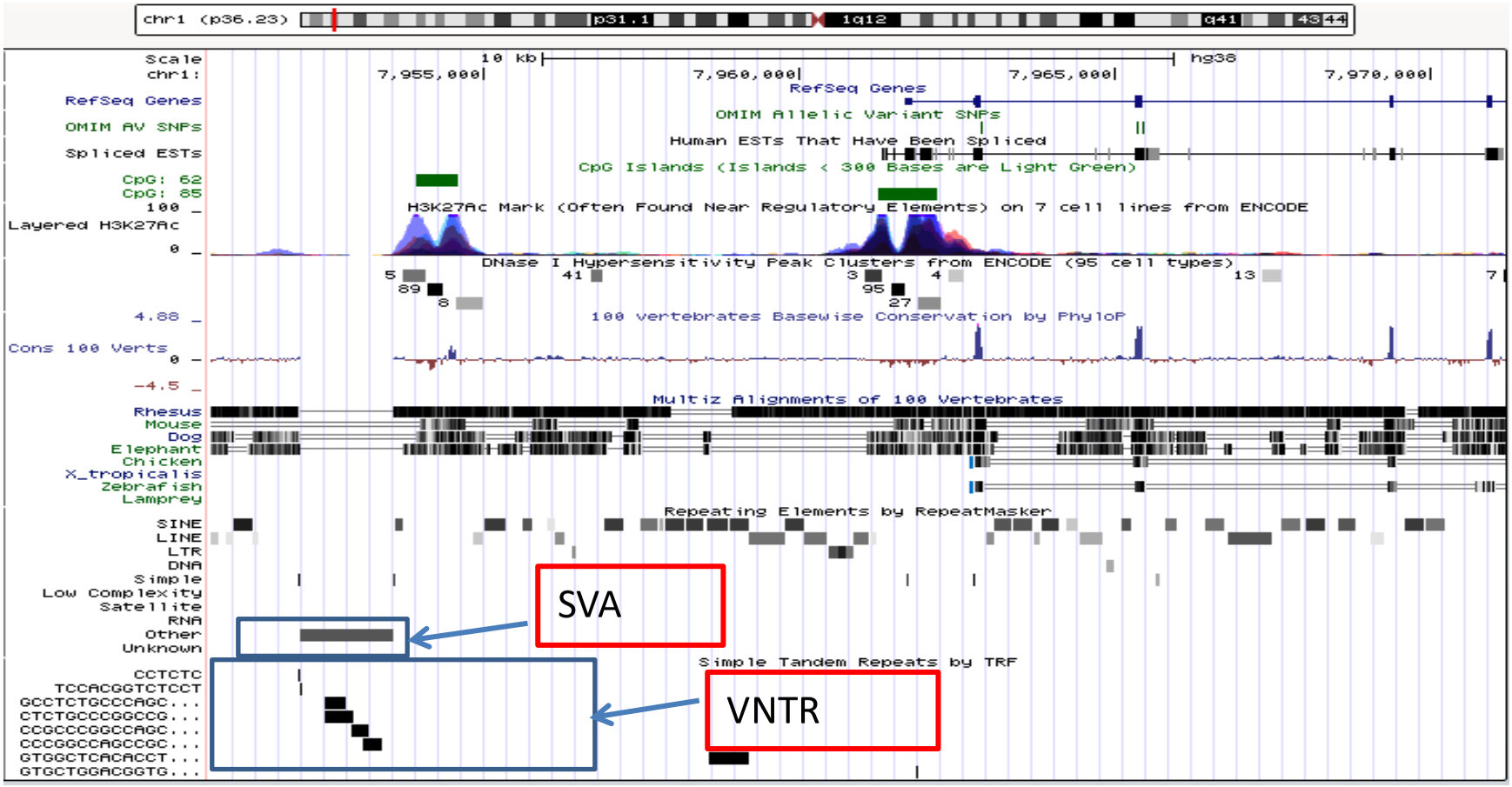


Figure 4.9: UCSC genome browser image upstream of the PARK7 gene. In the UCSC genomic browser, the location of the keys and tracks is located on the left hand side with scale and chromosomal loci present in the top of display. The illustrated region can be chosen by the user, in this screen shot showing, sequence SNPs, Human mRNAs, ESTs (expressed sequence tags), H3K27Ac histone mark (often found near regulatory elements), DNase clusters, transcription factor binding, overlapping the location of repetitive region SVA D and VNTR region sequences are boxed in blue mammalian conservation shared regions across species.

4.4.5 Endogenous PARK7 expression in E5 chick embryo brain

It was important to confirm the endogenous expression of the PARK7 gene in the chick brain at the development stage E5 since the analysis of reporter expression had to commence at this stage. This was confirmed by RT-PCR in the E5 chick brain. An expected band of 550 bp was obtained at 62.5 ºC annealing temperature by PARK7 primers using gradient PCR (temperature range 55 ºC to 65 ºC) (Figure 4.10).

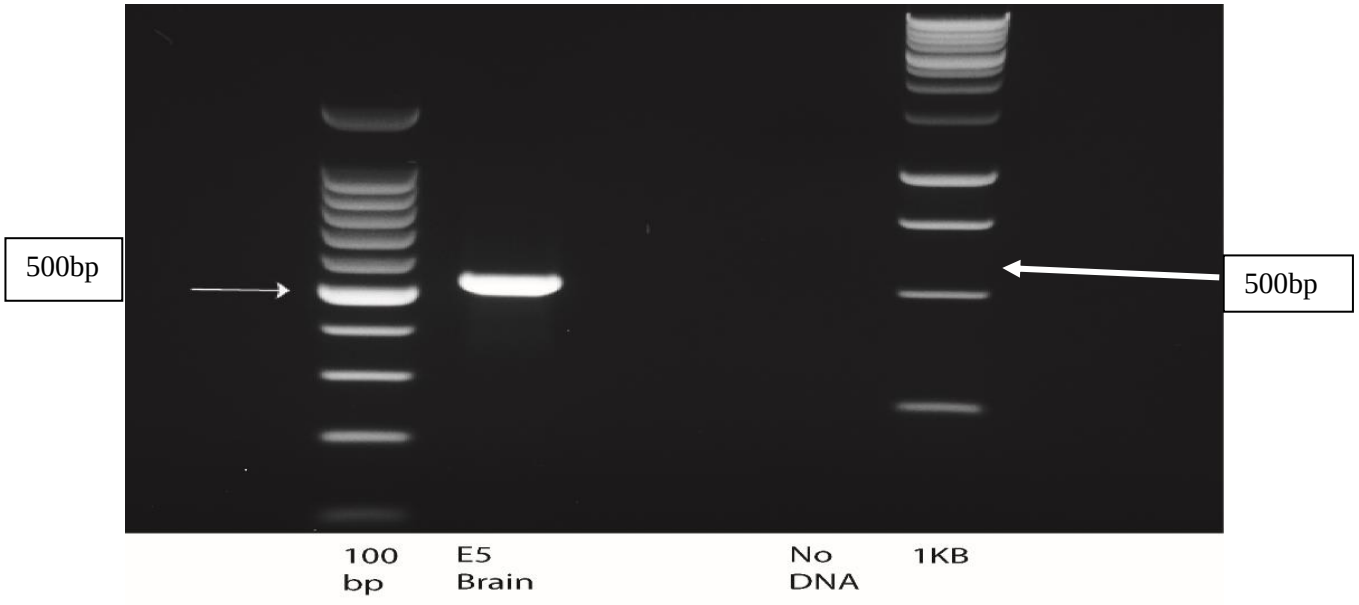


Figure 4.10 PARK7 is expressed in the brain tissue of E5 chick embryos.
Total RNA was purified from E5 brain reverse transcribed to cDNA and then used as a PCR template. 1.5% agarose gel showing a band at approximately the expected 550 bp. The MW marker is 100 bp with the white arrow pointing to 500 bp. The next lane was the negative control; the last lane was 1 kb ladder (Promega).

4.4.6 Generation of PARK7 SVA FUS PP phrGFP construct

The PARK7 SVA fragment was cloned into the FUS PP phrGFP plasmid, which was chosen to act as a minimal promoter. The original vector Zero Blunt PCR vector from Invitrogen) containing the PARK7 SVA fragment was a kind donation from Dr Abigail Savage. For the detail of the plasmid construction see Chapter 2, section 2.15.15.4 The PARK7 SVA fragment was released from the original vector PCR Blunt II by restriction enzyme digestion (Figure 4.11A). The PARK7 SVA was also run on the 0.8% gel electrophoresis and was shown to have a band size of 5 kb (Figure 4.11B). Meanwhile, the uncut FUS PP phrGFP and digested FUS PP phrGFP with *Nsil* were run on 0.8% gel electrophoresis, visualised by ethidium bromide nucleic stain and found to have a 4 kb band size (Figure 4.11C L1), digested FUSPP phrGFP ran lower than the uncut one (Figure 4.11C L2). The PARK7 SVA band at 1.6 kb was excised and cleaned over an SV column then dephosphorylated.

4.4 7 Ligation of the PARK7 SVA sequence into reporter vector FUS PP phrGFP to generate the PARK7 SVA FUS PP phrGFP

Ligation of the PARK7SVA into the FUSPPSVA phrGFP vector was performed. Selected colonies were mini-prepped and digested with Bgl1 for the presence of plasmid (Figure 4.11D). The PARK7 SVA FUS PP phrGFP was located in the forward orientation (Figure 4.11D L1), as the arrows indicate from the top at 2697 bp, 1724 bp, 1229 bp; and reverse orientation: 3075 bp, 1347 bp, 1229 bp (Figure 4.10D L2).The empty vector has two band at 2471bp, and 1229bp (Figure 4.11 (D) L3). The colonies with forward orientation were purified by plasmid prep, sequenced and then expanded in an overnight culture, PARK7SVA FUS PP phrGFP, was purified and concentrated to 11 µg/µl (Figure 6.19 Appendix).

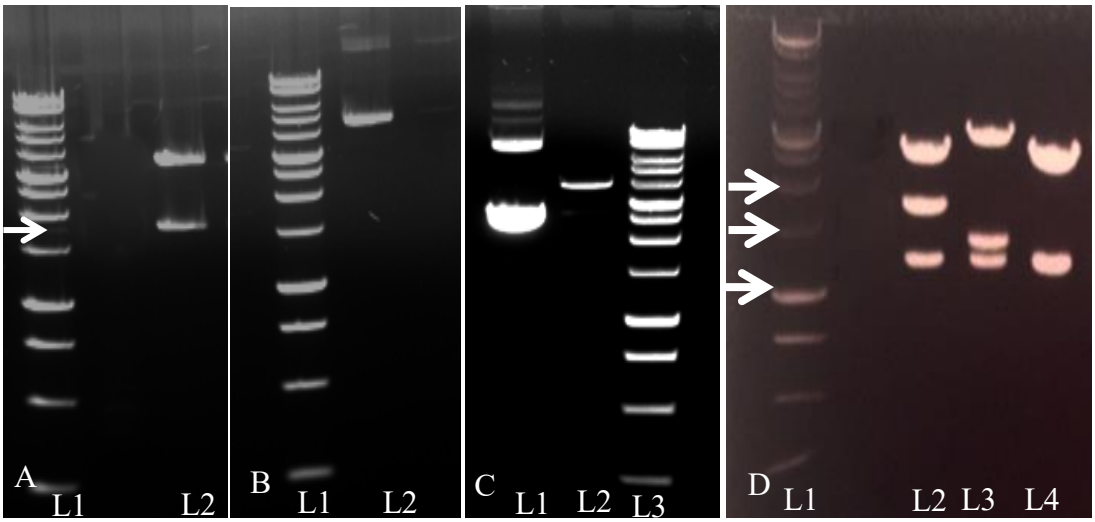


Figure 4.11: Steps of cloning PARK7 SVA into the reporter gene vector phrGFP. (A) L1 M W marker 1 kb, L2 cut of PARK7SVA with *Nsil* result band at 1.7 kb indicated by the white arrow. (B) L1 M W marker 1 kb, L2 uncut of PARK7SVA. (C) Digestion of phrGFP with *Nsil*. L1 shows a band of uncut of phr GFP. L2 show band of cut of phrGFP with *Nsil* result band at 4 kb. L3 M W marker 1 kb. (D) Digested minipreps with *BglI*. L1 M W marker 1 kb, L2 PARK7SVA in forward orientation, L3 PARK7SVA in reverse

4.4.8 The activity of the PARK7 SVA as a transcriptional regulator in reporter gene construct *in vitro*

The functional analysis of the PARK7 SVA *in vitro* was undertaken in the laboratory by Dr Abigail Savage and compiled in a published paper (Savage et al., 2013a). The data of the *in vitro* experiment are summarised here to demonstrate the functional activity of the PARK7 SVA in both SK-N-AS and MCF-7 cell lines. The PARK7 SVA in the forward orientation was cloned into the multiple cloning site of the pGL3P reporter gene vector upstream of the SV40 minimal promoter, when transfected into the SK-NA-S cell line no alteration in the level of expression was observed when compared with the minimal SV40 promoter alone. In the MCF-7 cell line, however, PARK7 SVA acted differently in the forward orientation. Here it significantly elevated the level of reporter gene expression. This indicated that intact PARK7 SVA responded differently in the two different cell types.

4.4.9 PARK7 SVA can act as a functional regulatory domain in an *in vivo* model

Injection and transfection of the PARK7SVA FUS PP phrGFP construct into chick neural tube**.**

The activity of the PARK7 SVA phrGFP construct in the forward orientation was tested in *vivo*. The neural tubes of chick embryos were injected and transfected by electroporation with an injection mixture containing PARK7 SVA at a final concentration of 4.2 µg, td Tomato 1 µg /µl and 1 µg/µl of Fast Green. After 48 hours of incubation, embryos which had been electroporated successfully in the neural tube, as evidenced by tomato plasmid expression, were tested for the PARK7 SVA PP phrGFP reporter gene by fluorescent microscopy. Three embryos were expressing tomato plasmid in the neural tube co-localised with GFP expression. The PARK7 SVA, therefore, exhibits reporter gene expression and demonstrates that PARK7 SVA has an activator property in chick embryo neural tubes compared with FUS SVA which acts as a weak activator (Figure 4.12) table 4.4.15.

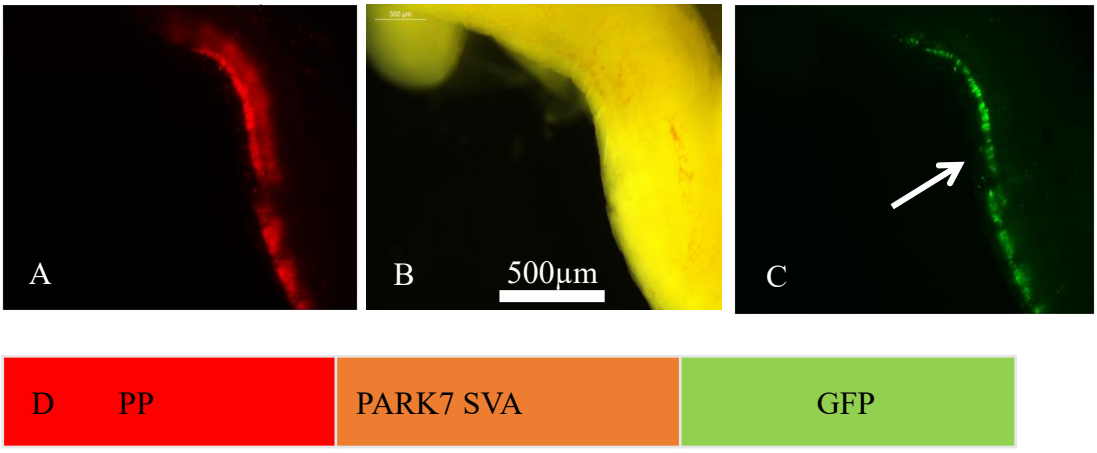


Figure 4.12: PARK7SVA FUSPP phrGFP activity in chick embryo neural tubes at E5. PARK7 SVA can enhance the expression of FUS PP. (A, B and C) show an embryo that has been electroporated with 1 µg/µl td tomato and 4.2 µg/µl PARK7SVA.(D) Schematic representation of PARK7 SVA phrGFP reporter gene vector

4.4.10 Develop a methodology to establish electroporation of chick brain with td Tomato at E5

Once it had been demonstrated that PARK7 had activity in the chick embryo neural tube at E5 it was decided to extend the investigation of the activity of the PARK7 SVA in the chick brain at E5. To obtain this data, the electroporation technique had to be modified to determine the optimum position and type of electrode. The linear wire electrode was substituted with a gold plated electrode. The position of the electrode was manipulated to obtain the best result. Different regions of the chick brain were transfected at HH stage 14, namely the hind brain, midbrain and forebrain. All three regions were transfected efficiently (data not shown), however the midbrain region was chosen for several reasons. First, the midbrain has a big cavity and can be injected very easily; second, the midbrain region at E2-3 consists of a thin layer of cells which allows exposure of a large area of cells to the plasmid mix meaning that many cells can be electroporated. Third, the midbrain can be positioned between the two electrodes and there is no tissue or blood vessels around to obstruct the electrode.

Initial experiments tested the td Tomato in the chick brain region as a positive control experiment. The midbrain or tectum region was injected with the td Tomato plasmid at HH stage 14. After 48 hours of incubation the chick embryos were observed under fluorescence microscopy for evidence of expression. The successfully transfected embryos showed strong red fluorescent protein at the site of injection and electroporation (Figure 4.13). Consequently, the midbrain region was used for transfection of the reporter gene construct and detection of the expression of the FUS PP alone, and the PARK7 FUS PP SVA PP, FUS PP SVA PP constructs, respectively.

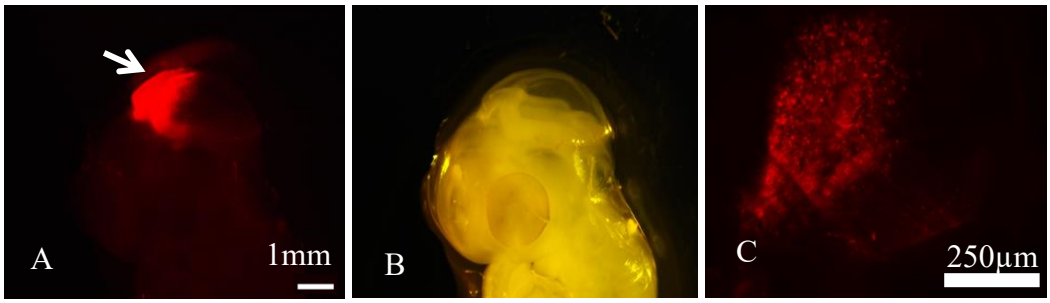


Figure 4.13: Chick embryo midbrain was transfected with td Tomato at E5. (A) Expression of td Tomato was detected in the midbrain region of chick embryos in vivo, indicated by the white arrow (B) bright field microscopy image. The midbrain region of a chick embryo was transfected with 1 µg /µl of tomato plasmid at HH 12-14. (C) A section from the transfected midbrain was dissected with a laboratory blade and placed under dissecting microscopy, showing the expression of tomato plasmid in the midbrain cells of the chick tectum neurones.

4.4.11 FUS PP phrGFP alone act as weak regulator in chick midbrain at E5

In the previous chapter the FUS PP phrGFP reporter gene construct was tested in the neural tube, where no expression activity was demonstrated. Now the FUS PP phrGFP plasmid in the chick embryo midbrain was tested to observe the activity of FUS PP alone in the brain tissue. A plasmid mix of 1 µg/µl of tomato plasmid and PBS with Fast Green and FUS PP phrGFP at a final concentration 4.4 µg was microinjected and then transfected by electroporation to the chick embryo midbrain at HH 12-14. The transfected embryos were incubated for 48 hours and viable embryos were analysed for td Tomato plasmid expression and then observed for GFP expression. All five embryos that had strong tomato expression also showed weak GFP expression in the midbrain; furthermore, the FUS PP was found to be active in the tectum cell sub type (Figure 4.14) table 4.4.15.

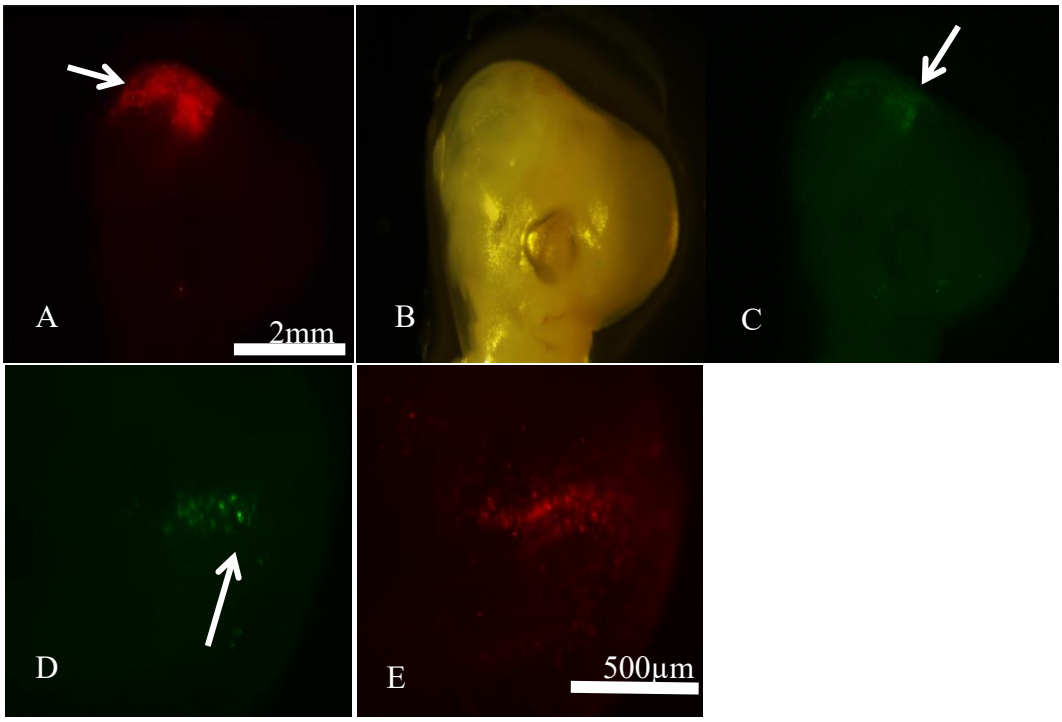


Figure 4.14: FUS PP phrGFP activity in the midbrain region of a chick embryo at E5. (A) td Tomato expression in chick midbrain, (B) bright field microscopy image, (C) FUS PP has weak expression in the midbrain, indicated by the white arrow. (D) Cross-section of the midbrain with weak FUS PP phrGFP expression observed in a subset of chick tectum cells, indicated by a white arrow. (E) The same section with td tomato expression.

4.4.12 PARK7 SVA acts as activator in the chick brain at E5

The PARK7 SVA was tested in the chick embryo midbrain. A mixture of the td Tomato at 1 µg/µl and the PARK7 SVA phrGFP at a final concentration 3.2 µg/µl was prepared and transfected into the cavity of the midbrain and then the embryos were incubated for 48 hours at 38 ºC in order to allow the reporter gene expression to develop (Figure 4.15) table 4.4.15. The PARK7 SVA FUS PP construct served as an activator and enhanced the expression of FUS PP in chick brain embryos, whereas in the previous experiment FUS PP alone was shown to have only a weak expression in the chick midbrain. Strong td tomato and GFP expression was observed in five embryos under fluorescent microscopy. PARK 7 SVA can therefore demonstrate activator behaviour in chick brains and neural tubes *in vivo,* and in the human neuroblastoma SK-N-AS cell line *in vitro.*

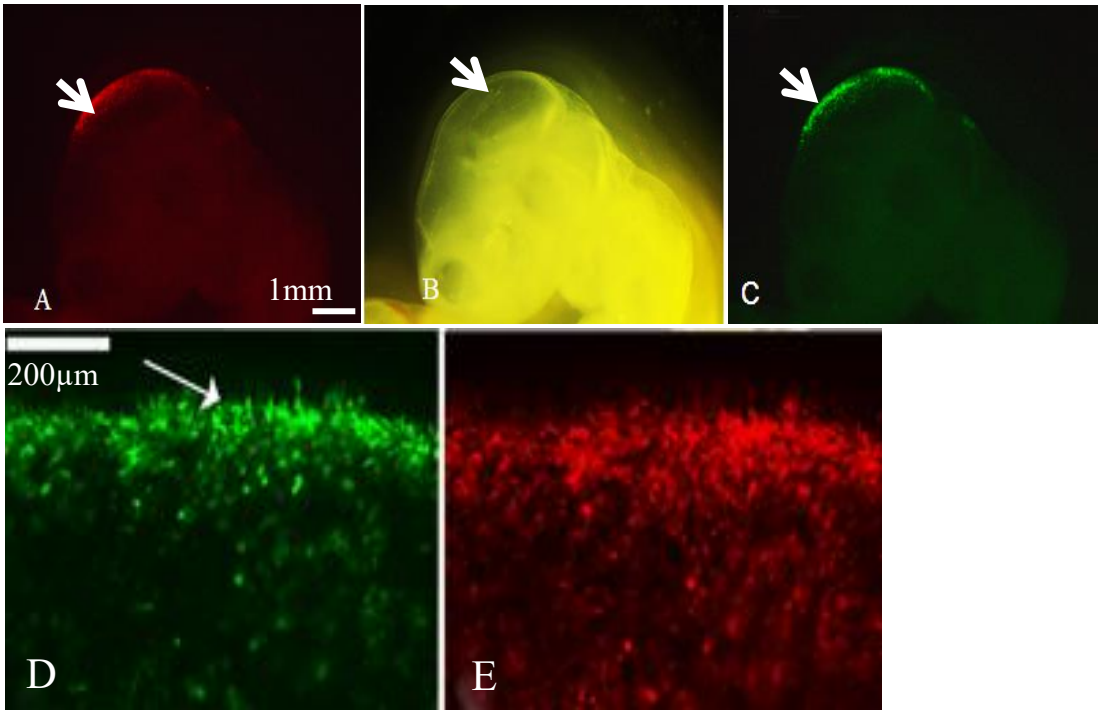


Figure 4.15: PARK7 SVA phrGFP activity in the midbrain region of a chick embryo at E5. (A) td Tomato expression, (B) bright field microscopy image, (C) PARK7SVA expression in the midbrain indicated by a white arrow. (D) Cross-section of the midbrain with GFP. (E) The same section with td tomato expression. (F) Schematic representation of PARK7 SVA phrGFP reporter gene vector

4.4.13 FUS SVA PP serves as an activator in chick brains at E5

FUS SVA PP phrGFP was tested in the chick embryo neural tube, to see if FUS SVA sequences can drive reporter gene expression in neural tubes. In this experiment the FUS SVA was tested in the midbrain region of chick embryo at E5. The plasmid mix of 1µg/µl of tomato plasmid and 1µl PBS with Fast Green and FUS SVA PP phrGFP at a final concentration 4.4 µg/µl was microinjected and then transfected by electroporation to the chick embryo midbrain at HH 12-14. Transfected embryos were incubated at 38 ºC for 48 hours before being tested for td Tomato expression and reporter gene expression of GFP. All three embryos that had td Tomato expression in the midbrain table 4.4.15 (Figure 4.16A) showed GFP expression in the same region (Figure 4.16C). Successful samples were fixed, then with a clean laboratory blade the transfected area of the midbrain was dissected (Figure 4.16B) and the site of the cut was indicated with a white line and white arrow, as in Figure 4.16 (D, E, and F). The dissected tissue from the midbrain was placed under a fluorescent microscope and it was evident that most tectum region cells have a strong td Tomato expression (Figure 4.16 (D). Strong FUS SVA expression was also observed under a GFP filter (Figure 4.16F), indicating the activation behaviour of the FUS SVA domain in the chick midbrain. Specifically, the expression of FUS SVA was found to be strong in a different subset of tectum cells than those that have td Tomato expression (Figure 4.16E).

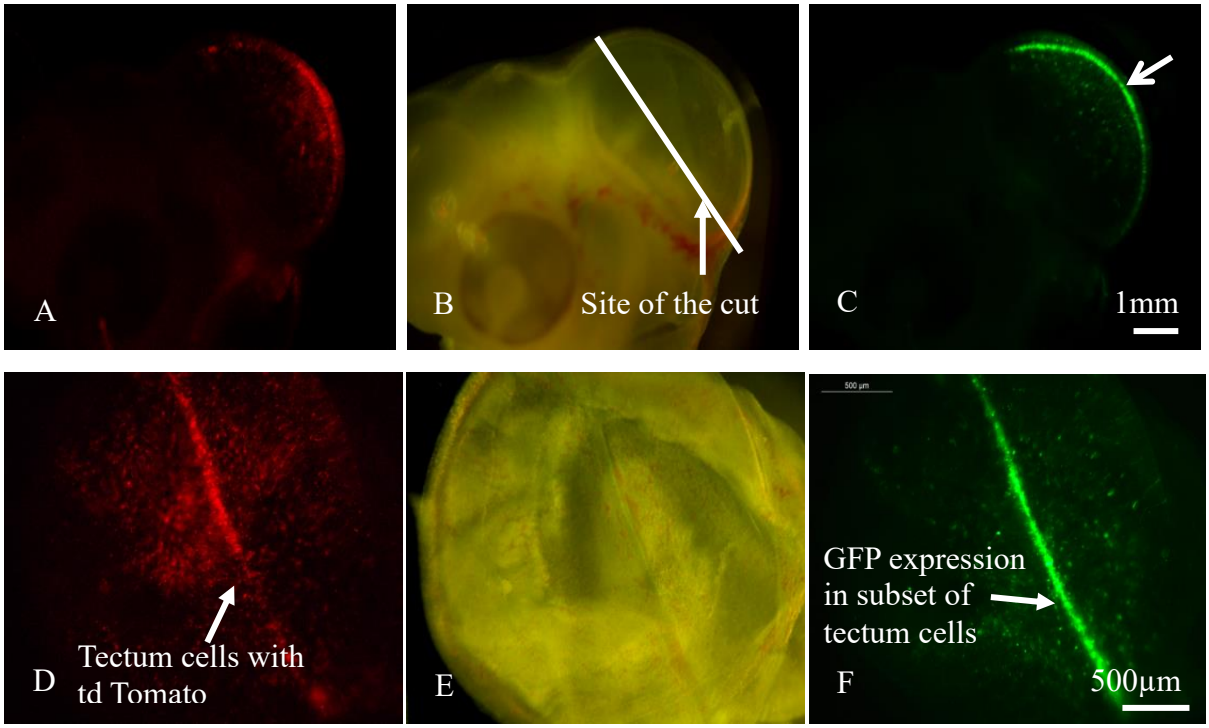


Figure 4.16: FUS SVA PP phrGFP activity in the midbrain region of a chick embryo at E5. (A) td Tomato expression, (B) bright field microscopy image, (C) FUS SVA PP has strong expression in the midbrain, indicated by the white arrow. (D) Cross-section of the midbrain with td Tomato (E) bright field microscopy image for the same section. (F) Cross-section with GFP expression.

4.4.14 Comparison of functional activity between FUS SVA, VNTR, PP and PARK7 SVA in the chick embryo central nervous system at E5

Comparing the overall functional activity of FUS SVA, FUS VNTR, and PARK7 SVA and FUSPP reporter Table 4.4.15, the results indicate that FUS SVA can act as weak enhancer in the neural tube since four embryos only show weak GFP expression in the neural tube out of 14 embryos with strong td Tomato expression. The same construct, however, has stronger enhancer activity in the chick brain. PRAK7 SVA, meanwhile, has enhanced activity in both the neural tube and chick brain. FUS PP alone is shown to have weak expression in the chick brain with no expression in the chick neural tube.

Table 4.4.15: Summary of findings (Location of injection (NT-neural tube, B-brain) HH -the Hamburger-Hamilton stage)

SVA, VNTR and FUS PP	Location of injection at stage 14 H.H.	H.H. Stage When harvested	td Tomato +ve Embryos	GFP +ve Embryos
F SVA PP	NT	22	14	4 weak GFP
F VNTR PP	NT		7	7
FUS SVA ECR PP	NT		6	6
FUS VNTR ECR PP	NT		3	3
PARK7 SVA PP	NT		3	3
PPARK7 SVA PP	B		5	5
FUS SVA PP	B		3	3
FUS PP	B		5	5 weak GFP
FUS PP	NT		10	0

4.5 Discussion

When the FUS gene locus was analysed for potential regulatory domains, an SVA retrotransposon was identified located 10 kb upstream of the MTSS of the gene. The FUS SVA, and its central VNTR, were tested *in vitro* and, interestingly, while the intact SVA was a repressor in the human neuroblastoma SK-N-AS cell line, the central VNTR region showed enhancer activity (Savage et al., 2014b). This data demonstrates the enhancer activity of the VNTR and the strong repressor element of the SVA that flanks this repetitive central VNTR region when analysed in the neuroblastoma cell line.

The SVA domain was then selected to be tested for its regulatory properties in the chick embryo neural tube and it was interesting to test a primate specific SVA D that was not conserved in chickens. From the previous results in Chapter 3 it was demonstrated that human FUS ECRs functioned very well in the chick embryo neural tube. This region of the neural tube contains the motor neurons, which are the appropriate cell type to test the regulatory properties that might be involved in ALS. To observe the regulatory properties of the FUS SVA this domain was first tested in the construct that works as a strong activator, namely the FUS ECR PP. When the SVA was cloned into the above plasmid and tested in the chick embryo neural tube for expression, the result showed that all embryos that have td Tomato expression also have GFP expression in the neural tube. Then the SVA was cloned in the FUS PP, it having already been shown in this research that this FUS PP when applied alone has no activity in the chick embryo neural tube. Fourteen embryos with the SVA FUS PP showed strong red fluorescence but only four out of these 14 embryos exhibited weak GFP expression in the neural tube. This demonstrates that this SVA functioned as weak activator for expression whereas a repressor activity was observed when SVA

was tested *in vitro* with the SV40 promoter. When the plasmid containing the FUS VNTR ECR PP was tested in the chick embryo neural tube, however, the transfected embryos showed strong red and green fluorescence in the neural tube and hence provided evidence for potential regulator activity.

In order to dissect further the activity of the components of the SVA, the VNTR was cloned in the FUS PP and tested in the chick embryo model. This showed that all transfected embryos were expressing td Tomato and GFP in the neural tube, indicating that VNTR has positive regulator properties. Since SVA has weaker activity than the VNTR this implies that there may be a repressor element in the SVA. These results reveal that there is more than one regulatory element in SVA perhaps an enhancer VNTR and a repressor in the remaining elements. This data suggests that there is a different balance of activity in human and chicks. This particular system of analysing the transcriptional properties of the domain is not quantitative, however, and therefore changes in the level of expression have not been determined here. Nonetheless, the fraction of transfected embryos that also fluoresce with GFP gives an indication of weak rather than strong expression. Furthermore, VNTR shows stronger activity in the chick neural tube, while the SVA containing VNTR is weaker, suggesting that an element within the SVA might have a repressor activity, and this partially matches what was seen *in vitro* in human cells. In the chick, the VNTR activation overwhelms the repression from the complete SVA elements, whereas the reverse is seen in human cells.

It was interesting to test the regulatory activity of another SVA located near the PARK 7 gene. This gene is known to be implicated in neurodegenerative diseases and breast cancer. This SVA is also of the subtype D and was located near to active histone markers and, according to the data obtained from both the UCSC Browser and ENCODE it might have regulatory properties. A previous study conducted by (Savage et al., 2013a) addressed human PARK7 SVA *in vitro* in SK-N-AS and breast cancer cell lines. The central VNTR from the PARK7 was analysed in the forward orientation, and showed different reporter gene expression in these two cell lines. The VNTR, when tested in the SK-N-AS, shows no activity when compared to the pGL3P SV40 minimal promoter. While SVA showed no activity in the SK-N-AS cell line it can enhance the reporter gene expression in MCF-7 cells (Savage et al., 2013a). This showed that SK-NA-S and breast cancer cells can respond differently to SVA and VNTR domain, undoubtedly because of the different cell environment; for example through transcription factors that get activated in the MCF-7 cells but not in the SK-N-AS ones (Savage et al., 2013b).

Functional analysis of the PARK7 SVA in the chick embryo model was conducted by selecting the PARK7 SVA in a forward orientation. This was cloned into the FUS PP phr GFP plasmid. The FUS PP plasmid was tested previously and showed no activity in chick embryo neural tubes, and this promoter can therefore act as a minimal promoter. When the plasmid PARK7 SVA FUS PP phrGFP was tested in chick embryo neural tubes the PARK7 SVA enhanced the reporter gene expression. Five embryos were shown to have GFP and td Tomato in the neural tube. After that, the function of the FUS SVA, VNTR and PARK7 SVA were tested in the chick embryo neural tube at E5. Then the investigation was extended into another region of central

nervous system; the chick brain. After many attempts to transfect the chick brain regions at stage HH12-14, it was shown that the midbrain or tectum can be transfected effectively with td Tomato alone. Strong red fluorescent can be observed under the fluorescent microscope at E5, indicating that this technique can be conducted successfully. A series of constructs were tested in chick brains at HH stage 12-14. The successfully transfected embryos were analysed for reporter activity at E5. The first plasmid tested was FUS PP alone, this plasmid was shown to have weak expression in chick embryo midbrains compared to no expression at all in chick neural tubes (see discussion above). PARK7 FUS PP was tested and shown to drive GFP expression in chick midbrains. Finally, FUS SVA PP also showed strong expression in chick midbrains and acts as an activator. In contrast, this plasmid was shown to act as a weak regulator in chick embryo neural tubes.

The work carried out in this chapter demonstrates the activity of two SVAs located upstream of two genes, FUS and PARK7, known for their implication in neurodegenerative disease and cancer. The activity of the SVA was previously addressed *in vitro* using human cell line neuroblastoma SK-N-AS and MCF-7 cells and here *in vivo* using chick embryo neural tubes and midbrains at E5. SVAs have regulatory ability *in vitro* and *in vivo,* which might depend on the cell phenotype and environment.

Chapter 5

General Discussion

98% of the human genome consists of DNA that does not code for proteins, and it is now becoming apparent that this should not be dismissed as 'junk' DNA, as it was formerly termed, but that these regions have domains that are vital in the regulation of gene expression. Gene expression is controlled in both spatial and temporal ways that operate at many levels; identification of new elements that regulate gene expression were the main aims of this project.

In the past identifying these regions was difficult for researchers, but the development of the Evolutionary Conserved Region (ECR) browser, the UCSC human genome browser and the utilisation of bioinformatics analysis now serve as useful tools in conjunction with the wealth of information available from the ENCODE project (much of which is on UCSC genome browser).

In the first part of this thesis, the ECR browser and UCSC genomic browser were used to predict important cis-acting DNA regulators of the FUS gene promoter that direct transcription (Khursheed et al., 2015). Putative regulatory domains were identified by cross-species comparison of the genomes; these domains were tested for function in reporter gene cassettes in the neural tube of a chick embryo model at E5 to address their potential as transcriptional regulatory domains. Two of the domains, located in the genome in intron 1 and a region upstream of the TSS of FUS directed reporter gene expression when they were adjacent to both the SV40 promoter and with the FUS proximal promoter, indicating that they can work with both heterologous and homologous promoters. The intronic domain INT1 is able to act as a transcriptional activator when it is cloned upstream of SV40 or phrGFP. Interestingly, INT1 shows the greatest activity in the hybrid neuroblastoma/ motor neurone cell line NSC-34, and could be involved in the expression of the FUS gene. The third domain (PP) identified in chapter 3, which encompassed the region Chr16: 31,191,429 had no intrinsic

activity in the chick neural tube at the stage of development analysed, and was used as a minimal promoter in later experiments in the neural tube. When (PP) domain was tested in the chick brain, however, it did show activity, suggesting there were essential differences in the cell environment or the cell phenotype enabling it to respond to the regulatory domain. Such differences could include transcription factors that were expressed in brain cells from those found in the motor neurone cells within neural tube.

In summary, ECRs were identified and can highlight important regulatory domains by using the ECR Browser and the UCSC genomic Browser. This approach can quickly highlight important putative regulatory domains of the human FUS gene based on *in vitro* reporter gene assays using comparative genomics of range of species such as chimpanzee, dog, rat, and mouse using 70% identity across a length of 100 bp as a parameter to identify the conserved regions in ECR browser. This analysis encompasses 4 kb upstream and downstream, coordinates of conserved regions were obtained from the UCSC genome browser (Hg19). Upon transcription factor analyses of the FUS gene using the ECR Browser linked rVista 2.0. The region spanning chr16:3119182-31192141 has 50 conserved transcription factors binding sites (TFBS) arranged into five groups, SP1, PAX5, E2F, E2F1, and E2F1DP1. Further analyses of the transcription factors coordinated to ch16:31191334-31191759 in the region of 426 bps showed 81 transcription factors were arranged in four groups, PAX5, SP1, E2F, and E2F1DP1. Further analyses of the transcription factors locates ch16:31191334-31191759 in the region of 426 bps showed 81 transcription factors, were arranged in four groups, PAX5, SP1, E2F, and E2F1DP1. Furthermore, ECRs in other genes often demonstrate important regulatory properties for example, studies have demonstrated the role of seven ECRs selected from MIR137 locus. MIR137

containing the micro RNA known for its significant associations with schizophrenia, these ECRs can regulate reporter expression in neuroblastoma cell line, SH-SY5Y (Gianfrancesco et al., 2016).

5.1 SVA domains upstream of the FUS and PARK7 gene play a role in regulation of gene expression

In the second part of this thesis another class / type domain was analysed, the SVA. This domain differs from an ECR by being primate specific, one of the most recent transposable elements to enter into the human genome. SVA SVA insertions have been identified in 96 diseases as of 2012 (Hancks and Kazazian, 2012). Furthermore, SVA insertions can have an impact in normal brain function and pathologically in tumours (Richardson et al., 2014). Retrotransposition in general can act as transcriptional or cause epigenetic modulation, specifically alterations in methylation patterns which may cause variation in cancer progression (Szpakowski et al., 2009). Many studies have demonstrated that the retrotransposons and SVAs lose their silencing property in cancer and the aging brain (Szpakowski et al., 2009, Baillie et al., 2011). Somatic retrotransposon has been demonstrated in the ageing brain and might play a role in ageing (Baillie et al., 2011, Faulkner, 2011). SVA has also been postulated to control the brain size in specific regions that are responsible for learning, muscle movement, balance and memory (Vasieva et al., 2016, Pulvers and Huttner, 2009). SVA insertions can have implications in human gender-related brain function through hormonal sensing; for example, SVA F insertion in the BRACA1 gene has been demonstrated to have a crosstalk function with oestrogen receptor ESR1(Gorrini et al., 2014, Kang et al., 2012) Furthermore, SVAs are present in genes that have been associated with a range of psychiatric conditions such as intellectual disability,

epilepsy and bipolar disorder (Cirlugea and O'Donohue, 2016). Interestingly, SVAs insertions have been suggested to act as impact factor on evolution and cognition in human. For example, the older SVA subtypes A-C has occurred in genes associated with primitive characteristics while SVA subtypes D-F1were occurred in genes involved with complex behavioural characteristics. SVA sub type F and F1 were found to occurred in genes that involved with immune system and accounted as a major modulator of CNS function, and psychiatric disorders for example schizophrenia, depression, and autism (Thakur et al., 2015). Therefore, SVAs were chosen from the FUS and PARK7 genes known for their association to neurodegenerative disease. Both of these SVAs are found approximately 10 kb, 8 kb upstream of the TSS of FUS and PARK7 genes. Although FUS SVA and Park7 SVA share similarity such as both being members of subtype D they also have differences: PARK7 SVA considered as human specific whereas FUS SVA is found in primate genome, PARK7 SVA has all domains of a canonical SVA while FUS SVA is missing both the CCCTCT repeat at its 5' end and the poly A tail at the 3' end. The ability of both FUS, PARK7 SVAs to affect gene expression were tested in a reporter gene vector, with a minimal promoter. FUS SVA showed to repress the expression in the SK-N-AS cell line (Khursheed et al., 2015). While intact PARK7 SVA, in forward orientation has no alteration in the level of the reporter gene activity when, tested in the SK-N-AS cell line, but it's can significantly increase the reporter gene activity in MCF-7 (human breast adenocarcinoma) cell lines (Savage et al., 2013a). Suggesting, that SVA encompass a repressor element such as SINE domain was acting as a repressor domain. The final aspect of the research involved exploring the regulatory function of the SVA further. The VNTRs domain and their ability to regulate gene expression has been, investigated in both FUS, and PARK7 genes. Which showed to be polymorphic in the

number of repeats of their VNTRs. Since a large tandem repeat domain was identified at 10 kb 5' of major transcriptional start sites of FUS genes using the UCSC genomic bioinformatic site (http:// genome.ucsc.edu). VNTRs domain within the FUS SVA was therefore chosen for further study. FUS VNTRs were increase the transcription activity of the reporter gene in the SK-N-AS cell line (Khursheed et al., 2015). Whereas, PARK7 VNTRs can demonstrated a repressor activity in neuroblastoma cell line SK-N-AS and in MCF-7 cells showed a similar activity to that observed with the minimal promoter alone (Savage et al., 2013a). Both, FUS SVA, PARK7 SVAs ability to affect gene expression were tested in an *in vivo* chick embryo model. FUS SVA, demonstrates as weak activator for expression in chick embryo neural tube (Figure 4.5) and as a strong activator in chick mid brain (Figure 4.16). PARK7 SVA in forward orientation enhanced the expression in chick embryo neural tube (Figure 4.12) and mid brain region (Figure 4.15). FUSVNTR shows to act as activator in chick embryo neural tube (Figure 4.7).

In summary our data suggests that SVA domains upstream of the FUS gene can potentially contribute to ALS by modulation of gene expression at the FUS gene locus and that variation of the VNTR can also regulate gene expression in accordance to the challenge that the cell receives.

Chick embryos have been used as a model system in this project to understand the role of the novel regulatory elements in regulating gene expression among genes important for their implication in motor neuron disease. This model system can interpret the results in a qualitative form and can define the repressor activity of the regulatory domains but they do not quantify the results since such quantification is dependent on several factors such as the efficiency of the electroporation, the sub set of transfected

cells, the development stage of the embryo and the rate of transfected cells. In order to use this system effectively, a program such as Image J software should be used to calculate the exact number of the transfected cells with internal control and with the reporter gene construct. Furthermore, the ratio of cells that are transfected with the control plasmid and the reporter plasmid can be employed. This model system has many advantages including accessibility, ease of tissue manipulation, and lack of cell death during electroporation, quick analysis just 48 hours after electroporation, comparable with *in vitro* methods. Our results have shown that the chick embryo model is useful and can be used to corroborate and extend *in vitro* reporter gene assays findings provided the gene of interest is expressed in the differentiating cells in the embryo. This model can work well particularly with comparative genome analysis where the human regulatory sequences are identified according to their conservation with chick or even fish genomes. To understand the role of regulatory regions on gene expression it is extremely valuable to employ efficient methods for gene transfer into living embryos because i) tissue architecture and the relationship between different cell types are maintained and ii) the normal physiological regulatory pathways and mechanisms are intact. Here we successfully used the application of an *in ovo* electroporation technique to analyse the regulation of gene expression in genes known for their implications in neurodegenerative disease.

5.2 Proposed Future experiments and directions

The dual fluorescent system used in this thesis, are qualitative and very subjective. This system cannot quantify the results for several reasons, for example the success of the electroporation, the development stage of the embryo, subset of transfected cells. In order to use this system effectively and capture the effect of the regulatory elements on gene expression, it would be important to employ methods to quantify the level of expression directed by the reporter gene. There are several approaches, such us, using dual fluorescence protein assay. In this system a marker protein (red fluorescence) can be driven by a constitutive promoter and compared with the amount of GFP directed by the test promoter (ECR/ VNTRs or SVA). The results can then be quantified using Image J software to measure the green: red fluorescence ratio.

Mutations or polymorphisms, within the cis-acting regulatory domains or 3' UTR mutations of FUS gene may have a major effect on the expression and potential dysregulation of FUS gene expression. Based on analysis of both FUS SVA and FUS VNTR reporter gene constructs *in vivo* chapter 4, we could assign differential regulatory properties to these regions *in vivo*; specifically, the long variant from FUS SVA has worked as a weak regulator, whereas when the FUS VNTR region of the long variant was as activator, indicating the multiple regulatory regions within the SVA as previously shown for the PARK7 SVA (Savage et al., 2013b). Conversely when comparing these same domain *in vitro*, it was the short variant that had higher activity in both cases. This would suggest distinct factors in both models are differentially modulating the activity, consistent with their ability to act as tissue specific and stimulus inducible manner.

Furthermore, based on analysis of the PARK7 SVA directed reporter gene expression when cloned in the forward orientation *in vivo* chapter 4, the PARK7 SVA was a positive regulator. In the *in vitro* model complete PARK7 SVA and its central VNTR can support differential reporter gene expression (Savage et al., 2013a). The PARK7 SVA, supports no reporter gene activity in SK-N-AS cells, but has activity in MCF-7 cells. Further dissection of the SVA, demonstrates that removal of the SINE element from PARK7 SVA, can increase the reporter gene expression, thus would be acting as a repressor domain. This would suggest testing regions from PARK7 SVA represented by VNTR, SINE, CCCTCT and Alu-like sequences individually, and study their distinct activity *in vivo*. The level of the PARK7 SVA reporter gene activity in the reverse orientation differs from PARK7 SVA activity in forward orientation *in vitro.* It would be useful to test the activity from PARK7 SVA in reverse orientation and study their regulatory function *in vivo*. The *in vivo* approach used in this thesis would be applicable to study several genes whose over expression is a key element in developing ALS, neurodegenerative disease, as it may give insight into regulation not observed in cell line models (Yamada et al., 2004).

Chapter 6

Appendix

6.1 Chicken primers were used for generating PARK7, FUS, ELP3, KIFAP3, and GAPDH genes.

Primer name	Sequence	Tm	GC %	Primer length	Product length	Annealing temperature
p[PARK7-Gg]fw	5'-ATGGCCTCGAAAAGAGCGTT-3'	60.32	50.00	20	533bp	55-65
p[PARK7-Gg]rv	5'-GCCACTTCCTTCCCCATCAG-3'	60.39	60.00	20		
p[FUS-Gg]fw	5'-CCAACGATTACAGCCAGACC-3'	60.52	55.0	20	994bp	57
p[FUS-Gg]rv	5'-GAAACCTTGATGGGGTTGC-3'	60.31	52.63	19		
p[ELP3-Gg]fw	5'-TTCCACTTAGCCAAGGATGC-3'	60.21	50.00	20	957bp	59
p[ELP3-Gg]rv	5'-GGCACAAACAAGTCCTTTGC-3'	60.68	50.00	20		
p[KIFAP3-Gg]fw	5'-GGTGGTTGAAGAATGCAAGC-3'	60.7	50	20	1kb	55
p[KIFAP3-Gg]rv	5'-GTGAGTTTGGGAAGCAGACC-3'	59.7	55	20		

p[GAPDH]fw	5'-TCCAAGTGGTGGCCATCAAT-3'	57.3	50.00	20	500bp	55
p[GAPDH]rv	5'-TTCTGGGCAGCACCTCTGTC-3'	61.4	60.00	20		

Table 6.1: Details of primers used in amplification of FUS, KIFAP3, ELP3, PARK7, and GAPDH. The sequence of the forward and reverse primers, melting temperatures, the primer length the product size, and a specific annealing temperature that used for each primer set.

6. 2 information regarding the injection and electroporation of chick embryo in the central nerve system CNS.

Stage of embryo at HH	Site of injection	Electroporation	Electrode	Conditions and plasmid mix
12-14 HH	Neural tube	Position the electrode on both side of neural tube	Straight wire electrode of 5 mm length	100 – 200 nl
12-14 HH	Mid brain	Position of the electrode on both side of embryo head precisely on both side of mid brain region	paddle electrode was attached to the anode and the straight wire electrode was attached to the cathode	Same

Table 6.2: Information regarding the injection and electroporation of the chick embryo in the neural tube and brain region. Embryo stage, site of the conditions and type of electrodes that have been used in the electroporation.

6. 3 Information regarding the plasmid used in transfection of chick embryo central nerve system.

FUS PP phr GFP	Neural tube
FUS ECR PP phrGFP	Neural tube
FUS PP int phrGFP	Neural tube
FUS ECRPP int phrGFP	Neural tube
FUS SVA PP phrGFP	Neural tube and midbrain
FUS SVA ECR PP phrGFP	Neural tube
FUS VNTR PP phrGFP	Neural tube
FUS VNTR ECR PP phrGFP	Neural tube
PARK7 SVA FUS PP phrGFP	Neural tube and midbrain
P1229 Hb9 enhancer	Neural tube

Table 6.3: The reporter gene constructs used *in vivo.* This table shows names of the reporter gene constructs that have been used *in vivo* and the electroporated regions of the central nerve system of chick embryo.

6. 4 Information regarding the antibodies used in immunofluorescence staining

Primary antibodies	Dilution used
anti FUS raised in rabbit	1:500,Abcam ab 23439
Hb9 raised in mouse	(1:5,Developmental Studies Hybridoma Bank, Department of Biology, University of Iowa).
Secondary antibodies	
Goat anti mouse Alexa 488	1:250 (Invitrogen)
Goat anti rabbit Alexa 488	1:250 (Invitrogen)
Goat anti mouse IgG Alexa 350	1:500 (Invitrogen)

Table 6. 4: Details of the antibodies used in immunofluorescence staining of neural tube cross section protocol. Company the antibody was obtained from and the amount used in the immunofluorescence staining.

6. 5 Information regarding the transfection of S K- NA-S cell line with FUS regulatory domains

FUS ECR constructs	Number of transfected SK-N-AS that has GFP expression for the reporter gene and tomato expression as a marker
Tomato plasmid act as positive control	47-50 cells transfected
FUS PP	10-11 cells transfected
FUS ECR PP	7-8 cells transfected
FUS INT PP	10-11 cells transfected
FUS INT ECR PP	16-17 cells transfected
phrGFP act as negative control	No transfected cells

Table 6.5: Information regarding the transfection of SK-NA-S cell line with FUS regulatory domains. FUS ECR constructs used in the cell transfection, number of transfected cells that has GFP expression for the reporter gene and tomato expression as a marker.

6.6 Alignment between predicted and sequenced INT PP, INT ECR PP, and PARK7 SVA

A INT PP

Score=1301 bits (704), Expect=0.0, Identities= 704/704 (100%), Gaps= 0/704

Strand=Plus/plus

```
Query 37        ATGGCCTCAAACGGTAGGTAAGGGCGCGAGGCGACGGCGGCGGCGCACCCGGCCGAGGCC 96
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509253  ATGGCCTCAAACGGTAGGTAAGGGCGCGAGGCGACGGCGGCGGCGCACCCGGCCGAGGCC 32509312

Query 97        TCCCAGCTGGGCTTTTCGTTTTCAGTGGGACCGGGGCGGCGATCCCGTGTGGGATTTTTT 156
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509313  TCCCAGCTGGGCTTTTCGTTTTCAGTGGGACCGGGGCGGCGATCCCGTGTGGGATTTTTT 32509372

Query 157       GGCGCCCCTGTGGCGGGAAGCCGCGGAGAAGAGTAACTGGAGGAGGCTGGTGTCGCCATT 216
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509373  GGCGCCCCTGTGGCGGGAAGCCGCGGAGAAGAGTAACTGGAGGAGGCTGGTGTCGCCATT 32509432

Query 217       TTGTTTCGCTCCTCTGGCCCTCGCGCGCGGGGCGGGAAGTCTTTTCTTTGCAGTCCGTTT 276
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509433  TTGTTTCGCTCCTCTGGCCCTCGCGCGCGGGGCGGGAAGTCTTTTCTTTGCAGTCCGTTT 32509492

Query 277       GCTTGGGGTGGGCGTTGGGAGGGACGCTTCTTAGGGGTTTGAAGCGTCAGGTGAGGGTGG 336
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509493  GCTTGGGGTGGGCGTTGGGAGGGACGCTTCTTAGGGGTTTGAAGCGTCAGGTGAGGGTGG 32509552

Query 337       AAAACGCCCATTCTCCGTGGCCTCGCCTCCCCCAACTCCCGGCCCCGCGCTCGAGCCCGC 396
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509553  AAAACGCCCATTCTCCGTGGCCTCGCCTCCCCCAACTCCCGGCCCCGCGCTCGAGCCCGC 32509612

Query 397       TTTGTCGCAGTGCTGCATCCGGGCACTCGCGGCGCGCACGCGCTCTGCGGGCCCCTCCCC 456
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509613  TTTGTCGCAGTGCTGCATCCGGGCACTCGCGGCGCGCACGCGCTCTGCGGGCCCCTCCCC 32509672

Query 457       CTTCGCGGCGCGGGTACCCCTTCCCCGCCTCGTGTTGGTTCAGCTTTCTGTCGCGAGACC 516
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509673  CTTCGCGGCGCGGGTACCCCTTCCCCGCCTCGTGTTGGTTCAGCTTTCTGTCGCGAGACC 32509732

Query 517       CTTCGCGGAAGACTCGGCGGCGCGCGTCCGGTGTGAGCCTTGTCCCTCAGTGGTCCTTCG 576
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509733  CTTCGCGGAAGACTCGGCGGCGCGCGTCCGGTGTGAGCCTTGTCCCTCAGTGGTCCTTCG 32509792

Query 577       CGAATGGGCGGGACCGCTCCGTTCCCGCCTGGGTTGCCACGCGGCTGGGGGCGGAGGCTC  636
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509793  CGAATGGGCGGGACCGCTCCGTTCCCGCCTGGGTTGCCACGCGGCTGGGGGCGGAGGCTC 32509852

Query 637       GGGATCGGGGCCGCCCTCTAGCTTAACGGTTTGGCGGCGGTGGTCAGGGTTCGACCAACG  696
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509853  GGGATCGGGGCCGCCCTCTAGCTTAACGGTTTGGCGGCGGTGGTCAGGGTTCGACCAACG 32509912

Query 697       GACTTGGGGACGGCCCGAGAGTTTTTCCCGCCTAAATTTCTTTC  740
                |||||||||||||||||||||||||||||||||||||||||||
Sbjct  32509913 GACTTGGGGACGGCCCGAGAGTTTTTCCCGCCTAAATTTCTTTC  32509956
```

Range 2: 32508764 to 32509192GenBankGraphics Next Match Previous Match First Match

Alignment statistics for match #2

Score	Expect	Identities	Gaps	Strand	Frame
641 bits(710)	0.0	397/429(93%)	5/429(1%)	Plus/Plus	

Features:

29691 bp at 5' side: polyserase-2 isoform 1 precursor61 bp at 3' side: RNA-binding protein FUS isoform 2

```
Query 786       GCCATCCTGGGTGAAAGCGGGGCCCAGCGAAGGGGCCCGGCCACAGGAATCTCGGTTCCA  845
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32508764  GCCATCCTGGGTGAAAGCGGGGCCCAGCGAAGGGGCCCGGCCACAGGAATCTCGGTTCCA 32508823

Query 846       CCCCGCTACTCCCGGCTGTGACTCCAGTTTCGTccccagccgccgggaccgcccccctcgc  905
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32508824  CCCCGCTACTCCCGGCTGTGACTCCAGTTTCGTCCCCAGCCGCCGGGACCGCCCCCTCGC 32508883

Query 906       cccgccccAGCGGGCACTCNN-CCGTACCACTGTGCCTTCATGGGGGTGGAGATAGATC  964
                ||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
```

```
Sbjct 32508884  CCCGCCCCCAGCGGGCACTCAGGCCGTACCACTGTGCCTTCATGGGGGTGGAGATAGATC 32508943

Query 965       GTGGGCTAGTCCTGCCGAGGAGAGAGGGGTTCTTCCTCAAAAAATATGATTATGTATAGT  1024
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32508944  GTGGGCTAGTCCTGCCGAGGAGAGAGGGGTTCTTCCTCAAAAAATATGATTATGTATAGT 32509003

Query 1025      ATTCGCATGATTCTAGTTAACTTGTTTCCCTTCTGCCTGCTCGGACCCTCTACCTGCCCT  1084
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509004  ATTCGCATGATTCTAGTTAACTTGTTTCCCTTCTGCCTGCTCGGACCCTCTACCTGCCCT 32509063

Query 1085      ACGAAGGGGGCGGAGTGCGTTCCTGCCTCCCCCTGCTCTT-CGCNTTTGGNGCGCGCCNG  1143
                |||||||||||||||||||||||||||||||||||||||| ||| ||||| ||||||| |
Sbjct 32509064  ACGAAGGGGGCGGAGTGCGTTCCTGCCTCCCCCTGCTCTTCCGCGTTTGGTGCGCGCCTG 32509123

Query 1144      CGCGGNNCGT--NNNNNGNANCGTACTTNAGCTTCGAC-NNNNNGNNNGGGNTGCTCANN  1200
                |||||  |||       | | ||||||| |||||||||      |   ||| ||||||
Sbjct 32509124  CGCGGTGCGTAGGCGGCGGAGCGTACTTAAGCTTCGACGCAGGAGGCGGGGCTGCTCAGT 32509183

Query  1201      CCTCCNGGC  1209
                 ||||| |||
Sbjct  32509184  CCTCCAGGC  32509192
```

B INT ECR PP

Score=1234 bites (1368) Expect 0.0

Identities	Gaps	Strand	Frame
684/684(100%)	0/684(0%)	Plus/Plus	

```
Query 37        ATGGCCTCAAACGGTAGGTAAGGGCGCGAGGCGACGGCGGCGGCGCACCCGGCCGAGGCC 96
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509253  ATGGCCTCAAACGGTAGGTAAGGGCGCGAGGCGACGGCGGCGGCGCACCCGGCCGAGGCC 32509312

Query 97        TCCCAGCTGGGCTTTTCGTTTTCAGTGGGACCGGGGCGGCGATCCCGTGTGGGATTTTTT 156
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509313  TCCCAGCTGGGCTTTTCGTTTTCAGTGGGACCGGGGCGGCGATCCCGTGTGGGATTTTTT 32509372

Query 157       GGCGCCCCTGTGGCGGGAAGCCGCGGAGAAGAGTAACTGGAGGAGGCTGGTGTCGCCATT 216
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509373  GGCGCCCCTGTGGCGGGAAGCCGCGGAGAAGAGTAACTGGAGGAGGCTGGTGTCGCCATT 32509432

Query 217       TTGTTTCGCTCCTCTGGCCCTCGCGCGCGGGGCGGGAAGTCTTTTCTTTGCAGTCCGTTT 276
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509433  TTGTTTCGCTCCTCTGGCCCTCGCGCGCGGGGCGGGAAGTCTTTTCTTTGCAGTCCGTTT 32509492

Query 277       GCTTGGGGTGGGCGTTGGGAGGGACGCTTCTTAGGGGTTTGAAGCGTCAGGTGAGGGTGG 336
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509493  GCTTGGGGTGGGCGTTGGGAGGGACGCTTCTTAGGGGTTTGAAGCGTCAGGTGAGGGTGG 32509552

Query 337       AAAACGCCCATTCTCCGTGGCCTCGCCTCCCCCAACTCCCGGCCCCGCGCTCGAGCCCGC 396
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509553  AAAACGCCCATTCTCCGTGGCCTCGCCTCCCCCAACTCCCGGCCCCGCGCTCGAGCCCGC 32509612

Query 397       TTTGTCGCAGTGCTGCATCCGGGCACTCGCGGCGCGCACGCGCTCTGCGGGCCCCTCCCC 456
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509613  TTTGTCGCAGTGCTGCATCCGGGCACTCGCGGCGCGCACGCGCTCTGCGGGCCCCTCCCC 32509672

Query 457       CTTCGCGGCGCGGGTACCCCTTCCCCGCCTCGTGTTGGTTCAGCTTTCTGTCGCGAGACC 516
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509673  CTTCGCGGCGCGGGTACCCCTTCCCCGCCTCGTGTTGGTTCAGCTTTCTGTCGCGAGACC 32509732

Query 517       CTTCGCGGAAGACTCGGCGGCGCGCGTCCGGTGTGAGCCTTGTCCCTCAGTGGTCCTTCG 576
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509733  CTTCGCGGAAGACTCGGCGGCGCGCGTCCGGTGTGAGCCTTGTCCCTCAGTGGTCCTTCG 32509792

Query 577       CGAATGGGCGGGACCGCTCCGTTCCCGCCTGGGTTGCCACGCGGCTGGGGGCGGAGGCTC 636
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509793  CGAATGGGCGGGACCGCTCCGTTCCCGCCTGGGTTGCCACGCGGCTGGGGGCGGAGGCTC 32509852

Query 637       GGGATCGGGGCCGCCCTCTAGCTTAACGGTTTGGCGGCGGTGGTCAGGGTTCGACCAACG 696
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32509853  GGGATCGGGGCCGCCCTCTAGCTTAACGGTTTGGCGGCGGTGGTCAGGGTTCGACCAACG 32509912

Query  697       GACTTGGGGACGGCCCGAGAGTTT  720
                  |||||||||||||||||||||||
Sbjct  32509913  GACTTGGGGACGGCCCGAGAGTTT  32509936
```

Score 652 bits(722) , Expect=0.0	Identities	Gaps	Strand	Frame
	403/435(93%)	5/435(1%)	Plus/Plus	

Features:
29685 bp at 5' side: polyserase-2 isoform 1 precursor61 bp at 3' side: RNA-binding protein FUS isoform 2

```
Query 767       GAGCTCGCCATCCTGGGTGAAAGCGGGGCCCAGCGAAGGGGCCCGGCCACAGGAATCTCG 826
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32508758  GAGCTCGCCATCCTGGGTGAAAGCGGGGCCCAGCGAAGGGGCCCGGCCACAGGAATCTCG 32508817

Query 827       GTTCCACCCCGCTACTCCCGGCTGTGACTCCAGTTTCGTccccagccgccgggaccgccc 886
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32508818  GTTCCACCCCGCTACTCCCGGCTGTGACTCCAGTTTCGTCCCCAGCCGCCGGGACCGCCC 32508877

Query 887       cctcgccccgcccccAGCGGGCACTCNN-CCGTACCACTGTGCCTTCATGGGGGTGGAGA 945
                |||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Sbjct 32508878  CCTCGCCCCGCCCCCAGCGGGCACTCAGGCCGTACCACTGTGCCTTCATGGGGGTGGAGA 32508937

Query 946       TAGATCGTGGGCTAGTCCTGCCGAGGAGAGAGGGGTTCTTCCTCAAAAAATATGATTATG 1005
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct 32508938  TAGATCGTGGGCTAGTCCTGCCGAGGAGAGAGGGGTTCTTCCTCAAAAAATATGATTATG 32508997

Query 1006      TATAGTATTCGCATGATTCTAGTTAACTTGTTTCCCTTCTGCCTGCTCGGACCCTCTACC 1065
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 32508998  TATAGTATTCGCATGATTCTAGTTAACTTGTTTCCCTTCTGCCTGCTCGGACCCTCTACC 32509057

Query 1066      TGCCCTACGAAGGGGGCGGAGTGCGTTCCTGCCTCCCCCTGCTCTT-CGCNTTTGGNGCG 1124
                |||||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |||
Sbjct 32509058  TGCCCTACGAAGGGGGCGGAGTGCGTTCCTGCCTCCCCCTGCTCTTCCGCGTTTGGTGCG 32509117

Query 1125      CGCCNGCGCGGNNCGT--NNNNNGNANCGTACTTNAGCTTCGAC-NNNNNGNNNGGGNTG 1181
                |||| ||||||  |||       | | ||||||| |||||||||      |   ||| ||
Sbjct 32509118  CGCCTGCGCGGTGCGTAGGCGGCGGAGCGTACTTAAGCTTCGACGCAGGAGGCGGGGCTG 32509177

Query 1182      CTCANNCCTCCNGGC  1196
                ||||  ||||| |||
Sbjct 32509178  CTCAGTCCTCCAGGC  32509192
```

C PARK7 SVA

	Identities	Gaps	Strand	Frame
Score=1738 bits(941), Expect=0.0	954/964(99%)	2/964(0%)	Plus/Plus	

Features:

```
Query  70       GGCTTTTTGATAACCCCTGACTGAAACCCTAAGTAATGGGATCTTACTCTCACTTCAAGA  129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7999801  GGCTTTTTGATAACCCCTGACTGAAACCCTAAGTAATGGGATCTTACTCTCACTTCAAGA  7999860

Query  130      AATAAGAtcctctccctctccctctccctctccctctccctctctctccacggtctcctt  189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7999861  AATAAGATCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCTCTCCACGGTCTCCTT  7999920

Query  190      ccacggtctccctctgatgccgagccaaagctggacggtactgctgccatctcggctcac  249
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7999921  CCACGGTCTCCCTCTGATGCCGAGCCAAAGCTGGACGGTACTGCTGCCATCTCGGCTCAC  7999980

Query  250      tgcaacctccctgcctgattctcctgcctcagcctgccgagtgcctgcgcacgccgccac  309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7999981  TGCAACCTCCCTGCCTGATTCTCCTGCCTCAGCCTGCCGAGTGCCTGCGCACGCCGCCAC  8000040

Query  310      gcctgactggttttcg--ttttttttttgtggagacggggttttgctgtgttggccgggc  367
                |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000041  GCCTGACTGGTTTTCGTTTTTTTTTTTTGTGGAGACGGGGTTTTGCTGTGTTGGCCGGGC  8000100

Query  368      tggtctccagctcctaaccacgagtgatccgccagcctcggcctcccgaggtgccgggat  427
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000101  TGGTCTCCAGCTCCTAACCACGAGTGATCCGCCAGCCTCGGCCTCCCGAGGTGCCGGGAT  8000160

Query  428      tgcagacggagtctcgttcactcagtgctcaatggtgcccaggctggagtgcagtggcgt  487
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000161  TGCAGACGGAGTCTCGTTCACTCAGTGCTCAATGGTGCCCAGGCTGGAGTGCAGTGGCGT  8000220

Query  488      gatctcggctcgctacaacctccacctcccagccgcctgccttggccccccaaagtgccg  547
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000221  GATCTCGGCTCGCTACAACCTCCACCTCCCAGCCGCCTGCCTTGGCCCCCCAAAGTGCCG  8000280

Query  548      agattgcagcctctgcccagccgccaccccgtctgggaagtgaggagcgtctctgcctgg  607
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000281  AGATTGCAGCCTCTGCCCAGCCGCCACCCCGTCTGGGAAGTGAGGAGCGTCTCTGCCTGG  8000340

Query  608      ccccccatcgtctgggatacgaggagcctctctgcctggctgcccagtctggaaagtgag  667
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000341  CCCCCCATCGTCTGGGATACGAGGAGCCTCTCTGCCTGGCTGCCCAGTCTGGAAAGTGAG  8000400

Query  668      gagcgtccctgcccggccgccatcccatctaggaagcgaggagcgcctcttccccgccgc  727
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000401  GAGCGTCCCTGCCCGGCCGCCATCCCATCTAGGAAGCGAGGAGCGCCTCTTCCCCGCCGC  8000460

Query  728      catcccatctaggaagtgaggagcgtctctgcccggccacccatcgtctgagatgtgggg  787
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000461  CATCCCATCTAGGAAGTGAGGAGCGTCTCTGCCCGGCCACCCATCGTCTGAGATGTGGGG  8000520

Query  788      agcacctctgccccgccgccctgtctgggatgtgaggagcgcctctgctgggccgcaacc  847
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8000521  AGCACCTCTGCCCCGCCGCCCTGTCTGGGATGTGAGGAGCGCCTCTGCTGGGCCGCAACC  8000580
```

```
Query  848      ctgtctgggaggtgaggagcgtctctgcccggccgccccgtctganaagtgagaaaaccc  907
                ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Sbjct  8000581  CTGTCTGGGAGGTGAGGAGCGTCTCTGCCCGGCCGCCCCGTCTGAGAAGTGAGAAAACCC  8000640

Query  908      tctgcctggcaaccgccccgtctganaantgaggancccctccgtccggcagccaccccg  967
                |||||||||||||||||||||||| || |||||| ||||||||||||||||||||||||
Sbjct  8000641  TCTGCCTGGCAACCGCCCCGTCTGAGAAGTGAGGAGCCCCTCCGTCCGGCAGCCACCCCG  8000700

Query  968      tctgggaantgaggancgtctccgcccggcagccaccccgtctgggganggaggtnggggg  1027
                |||||||| |||||| ||||||||||||||||||||||||||||||| |||||| |||||
Sbjct  8000701  TCTGGGAAGTGAGGAGCGTCTCCGCCCGGCAGCCACCCCGTCTGGGAGGGAGGTGGGGGG  8000760

Query  1028     gggg  1031
                ||||
Sbjct  8000761  GGGG  8000764
```

Figure 6.6:.Alignments between predicted and sequenced (from FUS proximal promoter) of (A) INT PP (B) INTECRPP (C) PARK7SVA sequences cloned into the phrGFP construct.

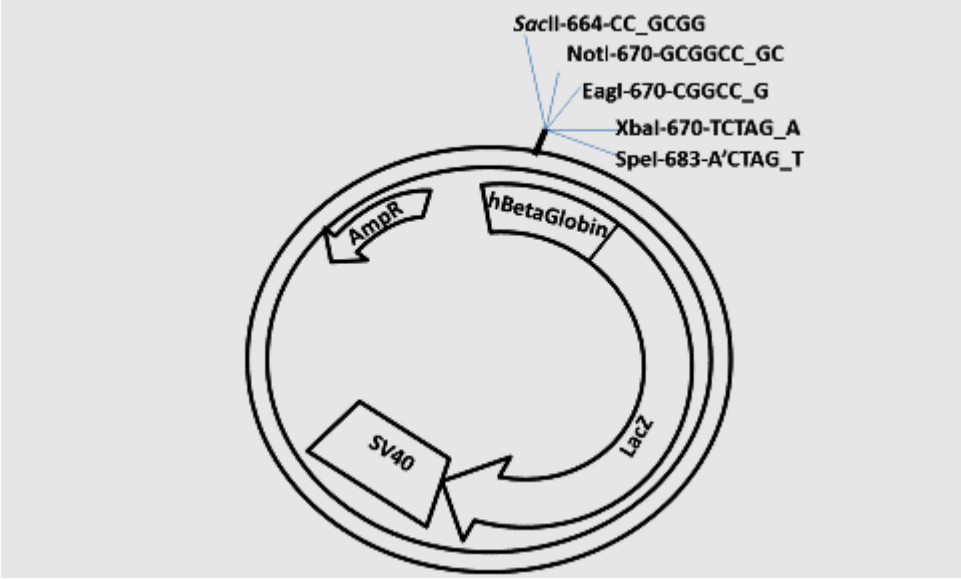


Figure 6.7: Map of P1229. The location of the Ampicillin resistance gene (AmpR) and minimal human β-globin promoter (hβglobin), LacZ gene and SV40 (Poly A generation) insertion into the pBluescript backbone. DNA sequence supplied by Alasdair Mackenzie. The RE site at the MCS that cut only once in the whole sequences are shown. Hb9 enhancer sequence was inserted by Dr Michael Lyons between NotI (670bp) and SpeI (+683bp).

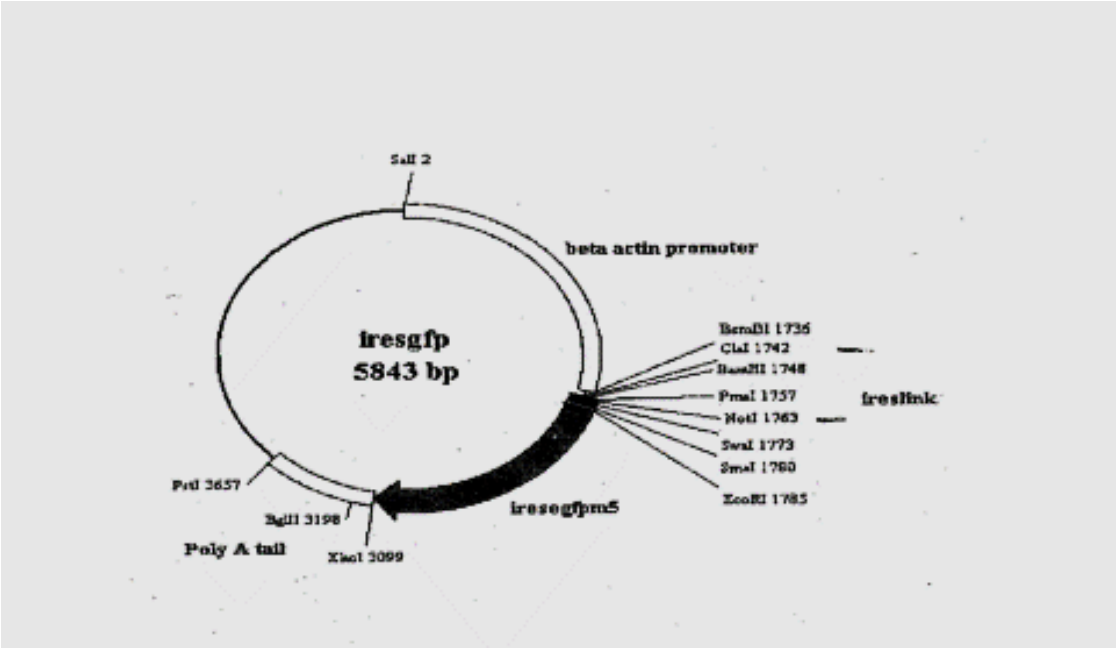


Figure 6.8: Map of ires GFP. The location of the minimal human β-actin Promoter DNA sequence, the RE site of the iresglink are shown. Ires GFP sequence inserted between XhoI (3099bp) and PstI (3657bp)

A7 Map of td Tomato

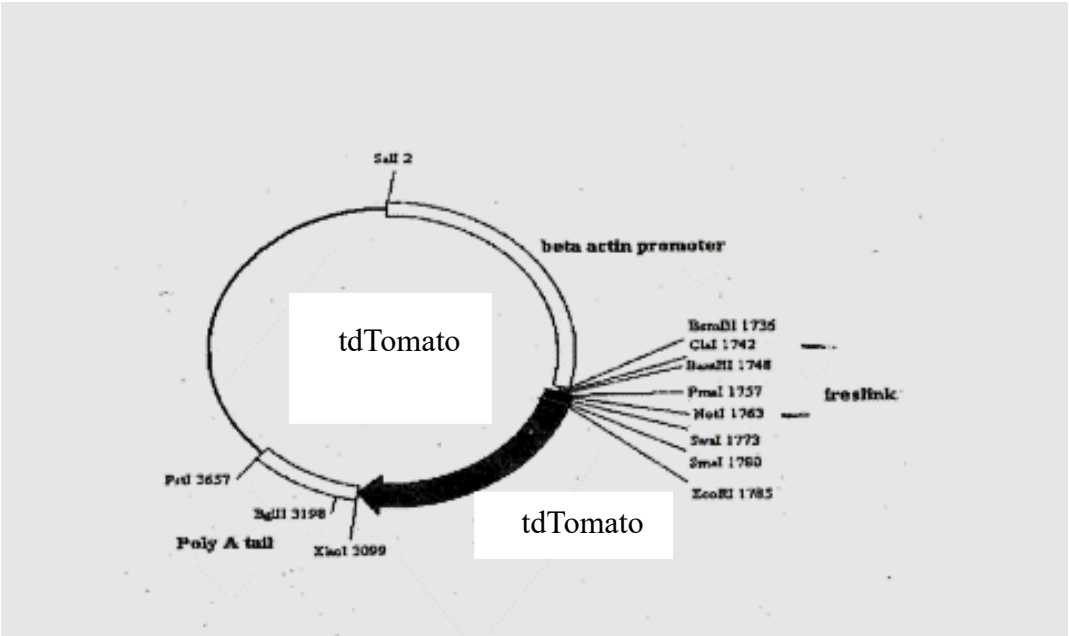


Figure 6.9: Map of td Tomato plasmid. The location of the β- actin promoter, DNA sequence of tdTomato a kind gift from Marco Marcello, University of Liverpool was inserted by Christine Cashman, University of Liverpool between EorRI (1785bp) and XhoI (3099bp)

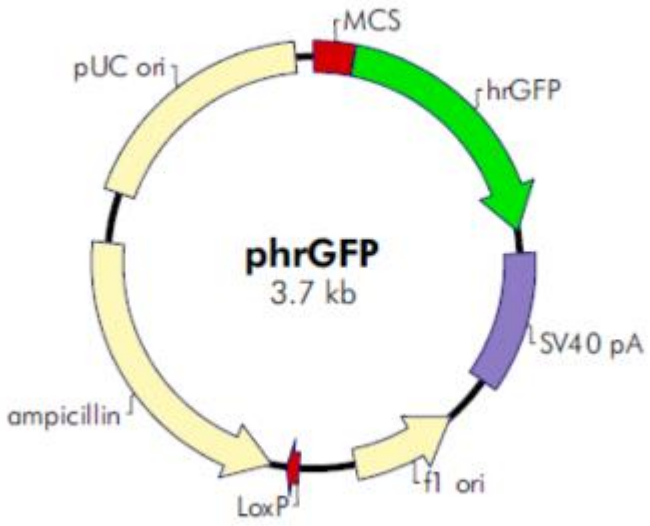


Figure 6.10: Map of the phrGFP Vector. This vector contains the hrGFP gene and the SV40 polyadenylation signal, multiple cloning site, the location of the Ampicillin resistance gene (AmpR) and lacks a eukaryotic promoter.

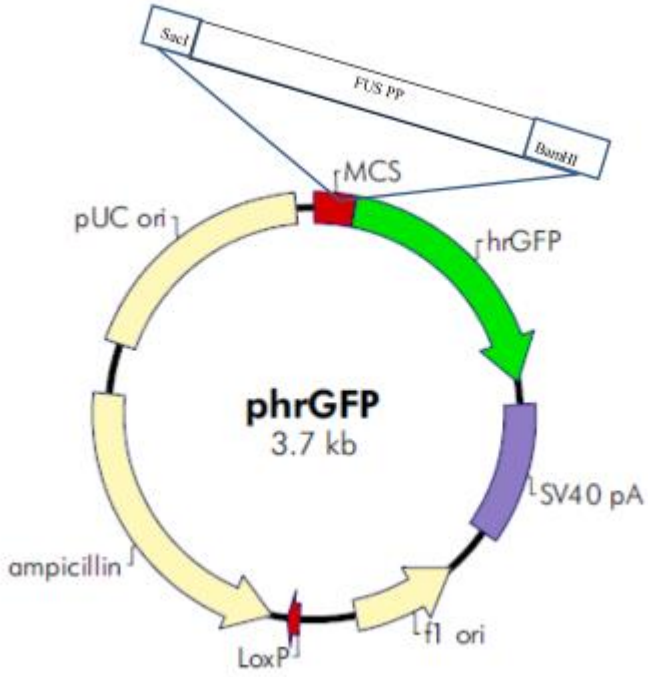


Figure 6.11: Map of phrGFP FUS PP the location of the Ampicillin resistance gene (AmpR) and minimal human FUS Promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS PP sequence was inserted between SacI (28bp) and BamHI (74bp).

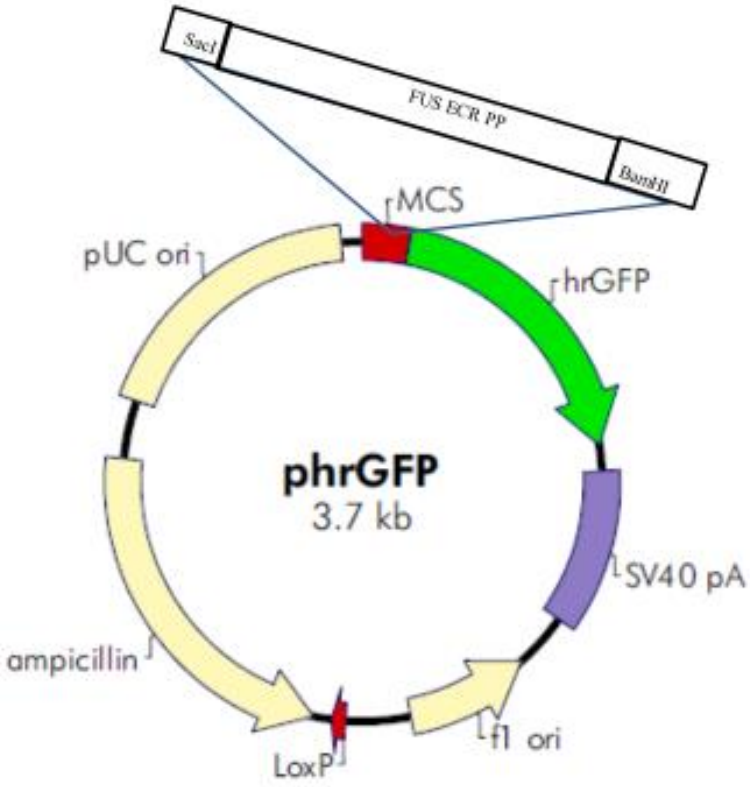


Figure 6.12: Map of phrGFP FUS ECR PP The location of the Ampicillin resistance gene (AmpR) and minimal human FUS Promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS PP sequence was inserted between SacI (28bp) and BamHI (74bp).

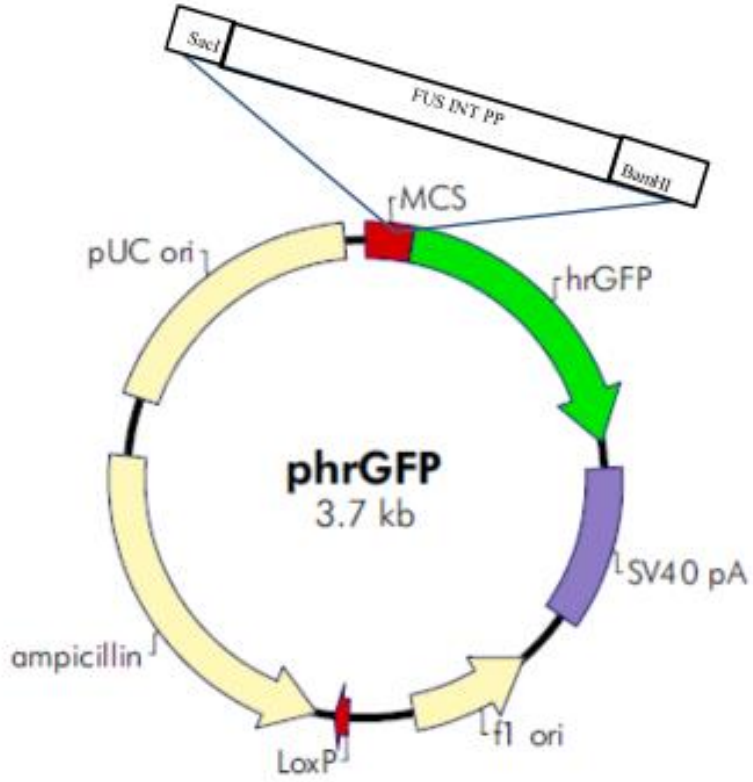


Figure 6.13: Map of phrGFP FUS INT PP The location of the Ampicillin resistance gene (AmpR) and minimal human FUS Promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS PP sequence was inserted between SacI (28bp) and BamHI (74bp).

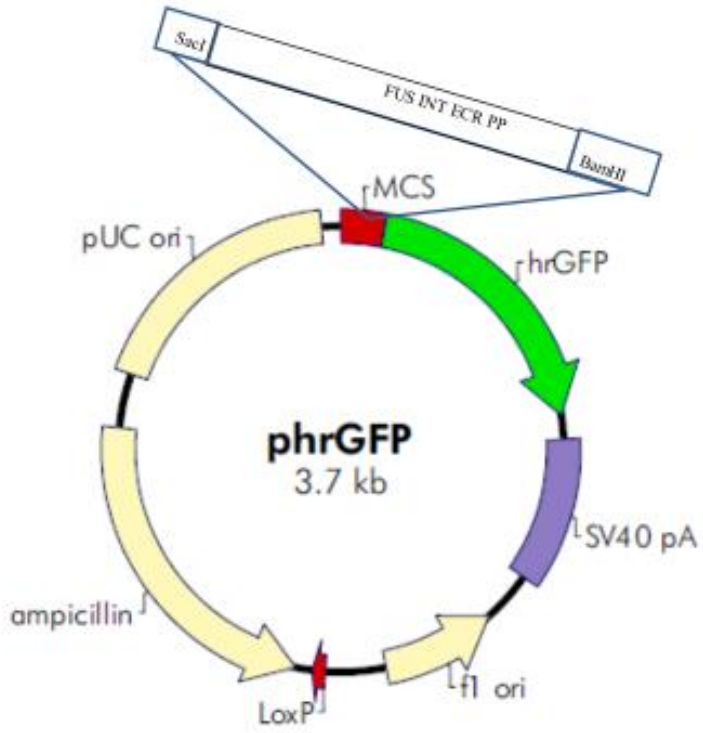


Figure 6.14: Map of phrGFP FUS INT ECR PP The location of the Ampicillin resistance gene (AmpR) and minimal human FUS Promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS PP sequence was inserted between SacI (28bp) and BamHI (74bp).

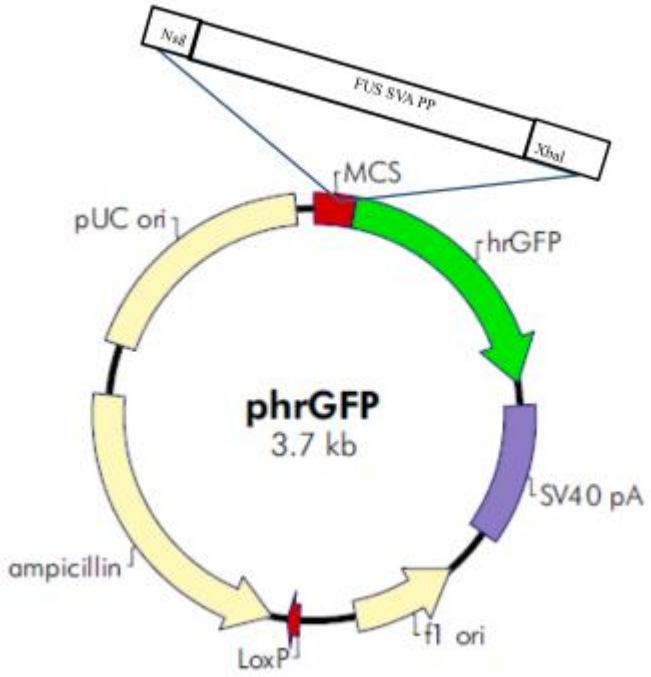


Figure 6.15: Map of phrGFP FUS SVA PP. The location of the Ampicillin resistance gene (AmpR) and minimal human FUS promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS SVA PP sequences was inserted between Nsil (1bp) and Xbal (13bp).

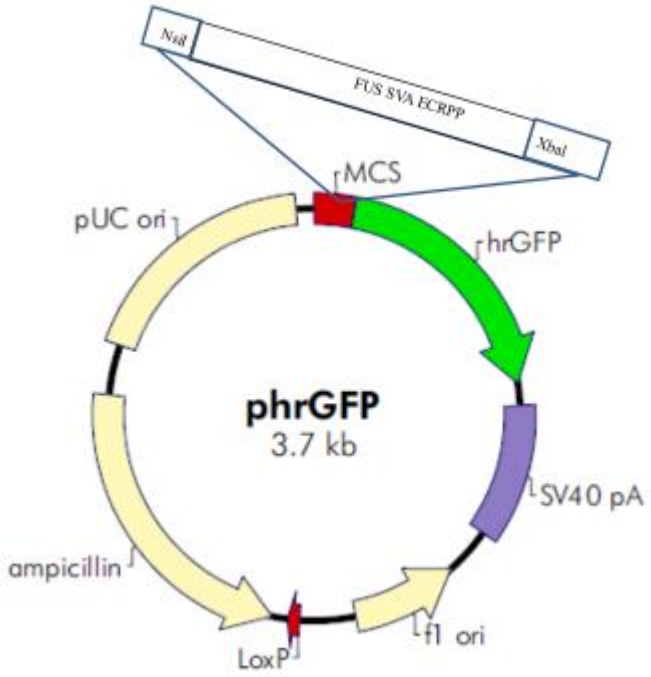


Figure 6.16: Map of phrGFP FUS SVA ECR PP. The location of the Ampicillin resistance gene (AmpR) and minimal human FUS promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS SVA PP sequences was inserted between Nsil (1bp) and Xbal (13bp).

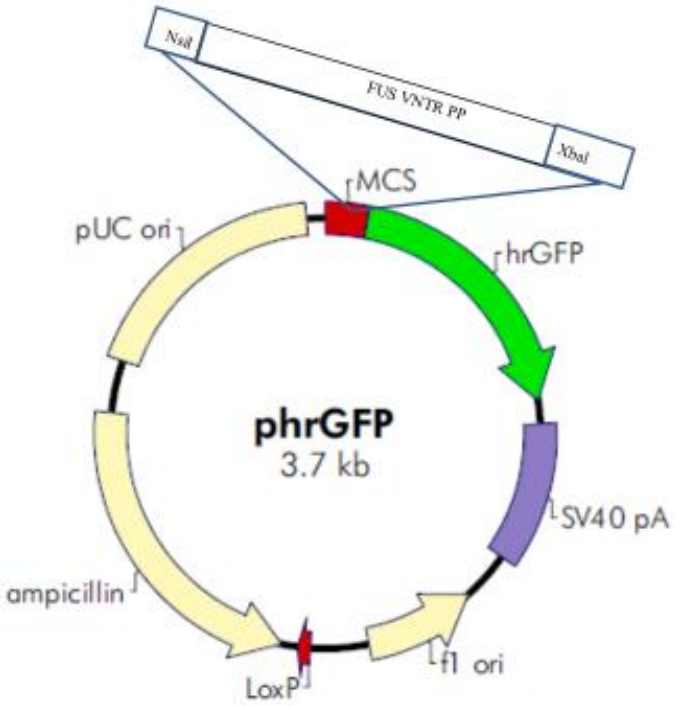


Figure 6.17: Map of phrGFP FUS VNTR PP. The location of the Ampicillin resistance gene (AmpR) and minimal human FUS promoter (FUS PP), GFP gene and SV40 (Poly A generation) insertion into the phrGFP backbone. FUS SVA PP sequences was inserted between Nsil (1bp) and Xbal (13bp).

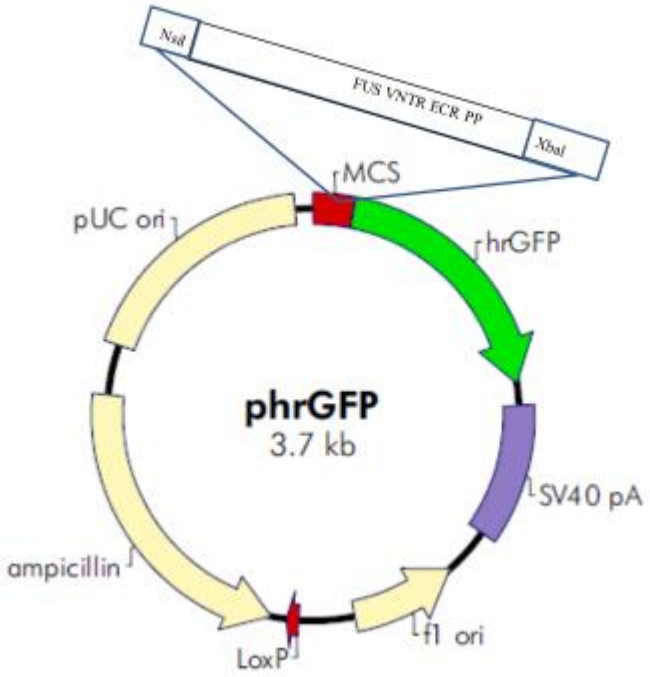


Figure 6.18: Map of phrGFP FUS VNTR ECR PP, the location of the Ampicillin resistance gene (AmpR) into the phrGFP backbone.FUS VNTR ECR PP sequence was inserted between Nsil (1bp) and Xba1 (13bp).

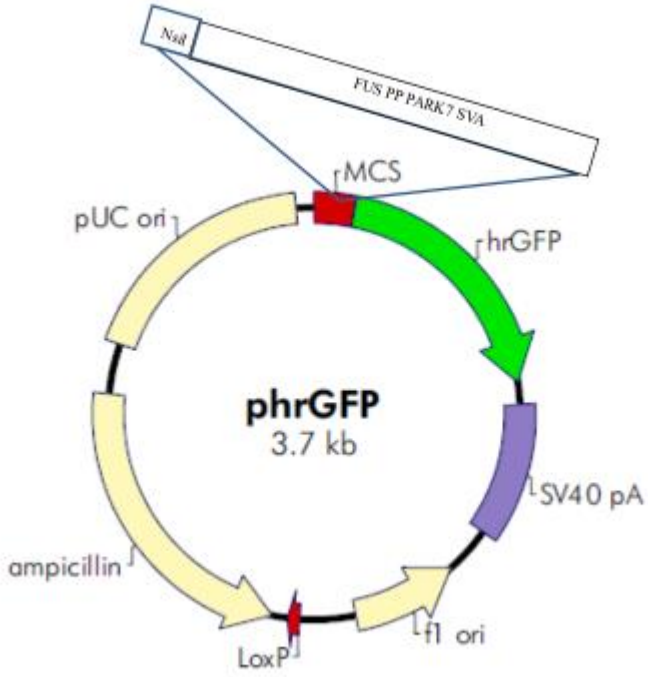


Figure 6.19: Map of phrGFP FUS PP PARK SVA, the location of the Ampicillin resistance gene (AmpR) into the phrGFP backbone.FUS VNTR ECR PP sequence was inserted in Nsil (1bp) site.

References

ABE, O., YAMADA, H., MASUTANI, Y., AOKI, S., KUNIMATSU, A., YAMASUE, H., FUKUDA, R., KASAI, K., HAYASHI, N. & MASUMOTO, T. 2004. Amyotrophic lateral sclerosis: diffusion tensor tractography and voxel-based analysis. *NMR in Biomedicine,* 17**,** 411-416.

AGOSTA, F., AL-CHALABI, A., FILIPPI, M., HARDIMAN, O., KAJI, R., MEININGER, V., NAKANO, I., SHAW, P., SHEFNER, J. & VAN DEN BERG, L. H. 2015. The El Escorial criteria: strengths and weaknesses. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration,* 16**,** 1-7.

AKHMEDOV, A. T., BERTRAND, P., CORTEGGIANI, E. & LOPEZ, B. S. 1995. Characterization of two nuclear mammalian homologous DNA-pairing activities that do not require associated exonuclease activity. *Proceedings of the National Academy of Sciences,* 92**,** 1729-1733.

AL-CHALABI, A., JONES, A., TROAKES, C., KING, A., AL-SARRAJ, S. & VAN DEN BERG, L. H. 2012. The genetics and neuropathology of amyotrophic lateral sclerosis. *Acta neuropathologica,* 124**,** 339-352.

ALVAREZ-CASTELAO, B., MUÑOZ, C., SÁNCHEZ, I., GOETHALS, M., VANDEKERCKHOVE, J. & CASTAÑO, J. G. 2012. Reduced protein stability of human DJ-1/PARK7 L166P, linked to autosomal recessive Parkinson disease, is due to direct endoproteolytic cleavage by the proteasome. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1823**,** 524-533.

ÅMAN, P., PANAGOPOULOS, I., LASSEN, C., FIORETOS, T., MENCINGER, M., TORESSON, H., HÖGLUND, M., FORSTER, A., RABBITTS, T. H. & RON, D. 1996. Expression Patterns of the Human Sarcoma-Associated GenesFUSandEWSand the Genomic Structure ofFUS. *Genomics,* 37**,** 1-8.

AOUN, S. M., CONNORS, S. L., PRIDDIS, L., BREEN, L. J. & COLYER, S. 2012. Motor Neurone Disease family carers' experiences of caring, palliative care and bereavement: an exploratory qualitative study. *Palliative medicine,* 26**,** 842-850.

ARBER, S., HAN, B., MENDELSOHN, M., SMITH, M., JESSELL, T. M. & SOCKANATHAN, S. 1999. Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron,* 23**,** 659-674.

BAILLIE, J. K., BARNETT, M. W., UPTON, K. R., GERHARDT, D. J., RICHMOND, T. A., DE SAPIO, F., BRENNAN, P. M., RIZZU, P., SMITH, S. & FELL, M. 2011. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature,* 479**,** 534-537.

BATZER, M. A. & DEININGER, P. L. 2002. Alu repeats and human genomic diversity. *Nature Reviews Genetics,* 3**,** 370-379.

BÄUMER, D., TALBOT, K. & TURNER, M. R. 2014. Advances in motor neurone disease. *Journal of the Royal Society of Medicine,* 107**,** 14-21.

BECK, C. R., GARCIA-PEREZ, J. L., BADGE, R. M. & MORAN, J. V. 2011. LINE-1 elements in structural variation and disease. *Annual review of genomics and human genetics,* 12**,** 187.

BECKMAN, J. S., CARSON, M., SMITH, C. D. & KOPPENOL, W. H. 1993. ALS, SOD and peroxynitrite. *Nature,* 364**,** 584-584.

BERTOLOTTI, A., LUTZ, Y., HEARD, D. J., CHAMBON, P. & TORA, L. 1996. hTAF (II) 68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *The EMBO journal,* 15**,** 5022.

BERTOLOTTI, A., MELOT, T., ACKER, J., VIGNERON, M., DELATTRE, O. & TORA, L. 1998. EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Molecular and cellular biology,* 18**,** 1489-1497.

BIER, E. & MCGINNIS, W. 2004. Model organisms in the study of development and disease. *Oxford Monographs On Medical Genetics.*

BOXER, A. L., MACKENZIE, I. R., BOEVE, B. F., BAKER, M., SEELEY, W. W., CROOK, R., FELDMAN, H., HSIUNG, G.-Y. R., RUTHERFORD, N. & LALUZ, V. 2011. Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *Journal of Neurology, Neurosurgery & Psychiatry,* 82**,** 196-203.

BREEN, G., COLLIER, D., CRAIG, I. & QUINN, J. 2008a. Engineering in Genomics [variable number tandem repeats as agents of functional regulation in the genome]. *IEEE Engineering in Medicine and Biology Magazine,* 27**,** 103-108.

BREEN, G., COLLIER, D., CRAIG, I. & QUINN, J. 2008b. Engineering in Genomics [variable number tandem repeats as agents of functional regulation in the genome]. *IEEE Engineering in Medicine and Biology Magazine,* 27.

BREEN, G., COLLIER, D., CRAIG, I. & QUINN, J. 2008c. Variable number tandem repeats as agents of functional regulation in the genome. *IEEE ENGINEERING IN MEDICINE AND BIOLOGY MAGAZINE,* 27**,** 103-+.

BROOKES, K.-J., MILL, J., GUINDALINI, C., CURRAN, S., XU, X., KNIGHT, J., CHEN, C.-K., HUANG, Y.-S., SETHNA, V. & TAYLOR, E. 2006. A common haplotype of the dopamine transporter gene associated with attention-deficit/hyperactivity disorder and interacting with maternal use of alcohol during pregnancy. *Archives of General Psychiatry,* 63**,** 74-81.

BUÉE-SCHERRER, V., BUEE, L., HOF, P. R., LEVEUGLE, B., GILLES, C., LOERZEL, A. J., PERL, D. P. & DELACOURTE, A. 1995. Neurofibrillary degeneration in amyotrophic lateral sclerosis/parkinsonism-dementia complex of Guam. Immunochemical characterization of tau proteins. *The American journal of pathology,* 146**,** 924.

BUHNIK-ROSENBLAU, K., DANIN-POLEG, Y. & KASHI, Y. 2012. Host genetics and gut microbiota. *Beneficial Microorganisms in Multicellular Life Forms.* Springer.

BURATTI, E., BRINDISI, A., GIOMBI, M., TISMINETZKY, S., AYALA, Y. M. & BARALLE, F. E. 2005. TDP-43 Binds Heterogeneous Nuclear Ribonucleoprotein A/B through Its C-terminal Tail AN IMPORTANT REGION FOR THE INHIBITION OF CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR EXON 9 SPLICING. *Journal of Biological Chemistry,* 280**,** 37572-37584.

BURKE, R. D., ATKINS, R. L. & WANG, D. 2005. Method and apparatus for targeting localized electroporation. Google Patents.

CALVIO, C., NEUBAUER, G., MANN, M. & LAMOND, A. 1995. Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *Rna,* 1**,** 724.

CASPI, A., SUGDEN, K., MOFFITT, T. E., TAYLOR, A., CRAIG, I. W., HARRINGTON, H., MCCLAY, J., MILL, J., MARTIN, J. & BRAITHWAITE, A. 2003. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science,* 301**,** 386-389.

CERERO LAPIEDRA, R., MORENO LOPEZ, L. & ESPARZA GOMEZ, G. 2002. Progressive bulbar palsy: a case report diagnosed by lingual symptoms. *Journal of oral pathology & medicine,* 31**,** 277-279.

CHANSKY, H. A., HU, M., HICKSTEIN, D. D. & YANG, L. 2001. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer research,* 61**,** 3586-3590.

CHEN, Y.-X., KRULL, C. E. & RENEKER, L. W. 2004. Targeted gene expression in the chicken eye by in ovo electroporation. *Mol Vis,* 10**,** 874-883.

CHINWALLA, A. T., COOK, L. L., DELEHAUNTY, K. D., FEWELL, G. A., FULTON, L. A., FULTON, R. S., GRAVES, T. A., HILLIER, L. W., MARDIS, E. R. & MCPHERSON, J. D. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature,* 420**,** 520-562.

CHURCH, D. M., GOODSTADT, L., HILLIER, L. W., ZODY, M. C., GOLDSTEIN, S., SHE, X., BULT, C. J., AGARWALA, R., CHERRY, J. L. & DICUCCIO, M. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol,* 7**,** e1000112.

CIRLUGEA, O. & O'DONOHUE, W. T. 2016. Review of psychometrics of forensic interview protocols with children. *Forensic Interviews Regarding Child Sexual Abuse.* Springer.

COLOMBRITA, C., ZENNARO, E., FALLINI, C., WEBER, M., SOMMACAL, A., BURATTI, E., SILANI, V. & RATTI, A. 2009. TDP-43 is recruited to stress granules in conditions of oxidative insult. *Journal of neurochemistry,* 111**,** 1051-1061.

DARDALHON, V., NORAZ, N., BOYER, M., BAKKER, A. Q., POLLOK, K., REBOUISSOU, C., SPITS, E. & TAYLOR, N. 1999. Green fluorescent protein as a selectable marker of fibronectin-facilitated retroviral gene transfer in primary human T lymphocytes. *Human gene therapy,* 10**,** 5-14.

DAWSON, T. M. & DAWSON, V. L. 2003. Rare genetic mutations shed light on the pathogenesis of Parkinson disease. *The Journal of clinical investigation,* 111**,** 145-151.

DE LUCA, V., ZAI, G., THARMALINGAM, S., DE BARTOLOMEIS, A., WONG, G. & KENNEDY, J. L. 2006. Association study between the novel functional polymorphism of the serotonin transporter gene and suicidal behaviour in schizophrenia. *European Neuropsychopharmacology,* 16**,** 268-271.

DENG, H. X., ZHAI, H., BIGIO, E. H., YAN, J., FECTO, F., AJROUD, K., MISHRA, M., AJROUD-DRISS, S., HELLER, S. & SUFIT, R. 2010. FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. *Annals of neurology,* 67**,** 739-748.

DILLMAN III, J. F., DABNEY, L. P., KARKI, S., PASCHAL, B. M., HOLZBAUR, E. L. & PFISTER, K. K. 1996. Functional analysis of dynactin and cytoplasmic dynein in slow axonal transport. *The Journal of neuroscience,* 16**,** 6742-6752.

DING, W., LIN, L., CHEN, B. & DAI, J. 2006. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB life,* 58**,** 677-685.

DION, P. A., DAOUD, H. & ROULEAU, G. A. 2009. Genetics of motor neuron disorders: new insights into pathogenic mechanisms. *Nature Reviews Genetics,* 10**,** 769-782.

DOOLITTLE, W. F. 2013. Is junk DNA bunk? A critique of ENCODE. *Proc Natl Acad Sci U S A.*

DRAKE, J. A., BIRD, C., NEMESH, J., THOMAS, D. J., NEWTON-CHEH, C., REYMOND, A., EXCOFFIER, L., ATTAR, H., ANTONARAKIS, S. E. & DERMITZAKIS, E. T. 2006. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nature genetics,* 38**,** 223-227.

ERICSON, J., BRISCOE, J., RASHBASS, P., VAN HEYNINGEN, V. & JESSELL, T. Graded sonic hedgehog signaling and the specification of cell fate in the ventral neural tube. Cold Spring Harbor symposia on quantitative biology, 1997. Cold Spring Harbor Laboratory Press, 451-466.

ERICSON, J., THOR, S., EDLUND, T., JESSELL, T. M. & YAMADA, T. 1992. Early stages of motor neuron differentiation revealed by expression of homeobox gene Islet-1. *Science,* 256**,** 1555-1560.

FARLEY, E. K. 2013. Gene Transfer in Developing Chick Embryos: In Ovo Electroporation. *Neural Development: Methods and Protocols***,** 141-150.

FAULKNER, G. J. 2011. Retrotransposons: mobile and mutagenic from conception to death. *FEBS letters,* 585**,** 1589-1594.

FENG, Q., MORAN, J. V., KAZAZIAN, H. H. & BOEKE, J. D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell,* 87**,** 905-916.
FERREIRA, M. A., O'DONOVAN, M. C., MENG, Y. A., JONES, I. R., RUDERFER, D. M., JONES, L., FAN, J., KIROV, G., PERLIS, R. H. & GREEN, E. K. 2008. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature genetics,* 40**,** 1056-1058.
FINK, J. K. 2013. Hereditary spastic paraplegia: clinico-pathologic features and emerging molecular mechanisms. *Acta neuropathologica,* 126**,** 307-328.
FISHER, K. M., ZAAIMI, B., WILLIAMS, T. L., BAKER, S. N. & BAKER, M. R. 2012. Beta-band intermuscular coherence: a novel biomarker of upper motor neuron dysfunction in motor neuron disease. *Brain***,** aws150.
FORMAN, M. S., FARMER, J., JOHNSON, J. K., CLARK, C. M., ARNOLD, S. E., COSLETT, H., CHATTERJEE, A., HURTIG, H. I., KARLAWISH, J. H. & ROSEN, H. J. 2006. Frontotemporal dementia: clinicopathological correlations. *Annals of neurology,* 59**,** 952-962.
FREIBAUM, B. D., CHITTA, R. K., HIGH, A. A. & TAYLOR, J. P. 2010. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *Journal of proteome research,* 9**,** 1104-1120.
FUNAHASHI, J. I., OKAFUJI, T., OHUCHI, H., NOJI, S., TANAKA, H. & NAKAMURA, H. 1999. Role of Pax-5 in the regulation of a mid-hindbrain organizer's activity. *Development, growth & differentiation,* 41**,** 59-72.
GABRIEL, B. & TEISSIE, J. 1997. Direct observation in the millisecond time range of fluorescent molecule asymmetrical interaction with the electropermeabilized cell membrane. *Biophysical journal,* 73**,** 2630.
GAL, J., ZHANG, J., KWINTER, D. M., ZHAI, J., JIA, H., JIA, J. & ZHU, H. 2011. Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. *Neurobiology of aging,* 32**,** 2323. e27-2323. e40.
GIANFRANCESCO, O., GRIFFITHS, D., MYERS, P., COLLIER, D. A., BUBB, V. J. & QUINN, J. P. 2016. Identification and potential regulatory properties of evolutionary conserved regions (ECRs) at the schizophrenia-associated MIR137 locus. *Journal of Molecular Neuroscience,* 60**,** 239-247.
GIJSELINCK, I., VAN LANGENHOVE, T., VAN DER ZEE, J., SLEEGERS, K., PHILTJENS, S., KLEINBERGER, G., JANSSENS, J., BETTENS, K., VAN CAUWENBERGHE, C. & PERESON, S. 2012. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *The Lancet Neurology,* 11**,** 54-65.
GITCHO, M. A., BALOH, R. H., CHAKRAVERTY, S., MAYO, K., NORTON, J. B., LEVITCH, D., HATANPAA, K. J., WHITE, C. L., BIGIO, E. H. & CASELLI, R. 2008. TDP-43 A315T mutation in familial motor neuron disease. *Annals of neurology,* 63**,** 535-538.
GOLDSTEIN, L. S. & YANG, Z. 2000. Microtubule-based transport systems in neurons: the roles of kinesins and dyneins. *Annual review of neuroscience,* 23**,** 39-71.
GOODIER, J. L. & KAZAZIAN, H. H. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell,* 135**,** 23-35.
GÖRANSSON, M., WEDIN, M. & ÅMAN, P. 2002. Temperature-Dependent Localization of TLS–CHOP to Splicing Factor Compartments. *Experimental cell research,* 278**,** 125-132.
GORRINI, C., GANG, B. P., BASSI, C., WAKEHAM, A., BANIASADI, S. P., HAO, Z., LI, W. Y., CESCON, D. W., LI, Y.-T. & MOLYNEUX, S. 2014. Estrogen controls the survival of BRCA1-deficient cells via a PI3K–NRF2-regulated pathway. *Proceedings of the National Academy of Sciences,* 111**,** 4472-4477.
GRAD, L. I. & CASHMAN, N. R. 2014. Prion-like activity of Cu/Zn superoxide dismutase: implications for amyotrophic lateral sclerosis. *Prion,* 8**,** 33-41.

GREENWOOD, C., SELTH, L. A., DIRAC-SVEJSTRUP, A. B. & SVEJSTRUP, J. Q. 2009. An iron-sulfur cluster domain in Elp3 important for the structural integrity of elongator. *Journal of Biological Chemistry,* 284**,** 141-149.

GUINDALINI, C., HOWARD, M., HADDLEY, K., LARANJEIRA, R., COLLIER, D., AMMAR, N., CRAIG, I., O'GARA, C., BUBB, V. J. & GREENWOOD, T. 2006. A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 4552-4557.

HADDLEY, K., VASILIOU, A., ALI, F., PAREDES, U., BUBB, V. & QUINN, J. 2008. Molecular genetics of monoamine transporters: relevance to brain disorders. *Neurochemical research,* 33**,** 652-667.

HAFEZPARAST, M., KLOCKE, R., RUHRBERG, C., MARQUARDT, A., AHMAD-ANNUAR, A., BOWEN, S., LALLI, G., WITHERDEN, A. S., HUMMERICH, H. & NICHOLSON, S. 2003. Mutations in dynein link motor neuron degeneration to defects in retrograde transport. *Science,* 300**,** 808-812.

HALLIER, M., LERGA, A., BARNACHE, S., TAVITIAN, A. & MOREAU-GACHELIN, F. 1998. The transcription factor Spi-1/PU. 1 interacts with the potential splicing factor TLS. *Journal of Biological Chemistry,* 273**,** 4838-4842.

HANCKS, D. C. & KAZAZIAN, H. H. SVA retrotransposons: Evolution and genetic instability. Seminars in cancer biology, 2010. Elsevier, 234-245.

HANCKS, D. C. & KAZAZIAN, H. H. 2012. Active human retrotransposons: variation and disease. *Current opinion in genetics & development,* 22**,** 191-203.

HARDISON, R. C. 2000. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends in Genetics,* 16**,** 369-372.

HATTEN, M. E. 1999. Central nervous system neuronal migration. *Annual review of neuroscience,* 22**,** 511-539.

HEILS, A., MÖSSNER, R. & LESCH, K. 1997. The human serotonin transporter gene polymorphism-basic research and clinical implications. *Journal of neural transmission,* 104**,** 1005-1014.

HERMAN, A. I., KAISS, K. M., MA, R., PHILBECK, J. W., HASAN, A., DASTI, H. & DEPETRILLO, P. B. 2005. Serotonin transporter promoter polymorphism and monoamine oxidase type A VNTR allelic variants together influence alcohol binge drinking risk in young women. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 133**,** 74-78.

HIBINO, M., ITOH, H. & KINOSITA JR, K. 1993. Time courses of cell electroporation as revealed by submicrosecond imaging of transmembrane potential. *Biophysical journal,* 64**,** 1789.

HILLIER, L. W., MILLER, W., BIRNEY, E., WARREN, W., HARDISON, R. C., PONTING, C. P., BORK, P., BURT, D. W., GROENEN, M. A. & DELANY, M. E. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature,* 432**,** 695-716.

HIROKAWA, N. & TAKEMURA, R. 2004. Kinesin superfamily proteins and their various functions and dynamics. *Experimental cell research,* 301**,** 50-59.

HOF, P. R., NIMCHINSKY, E., BUEE-SCHERRER, V., BUÉE, L., NASRALLAH, J., HOTTINGER, A., PUROHIT, D., LOERZEL, A., STEELE, J. & DELACOURTE, A. 1994. Amyotrophic lateral sclerosis/parkinsonism-dementia complex of Guam: quantitative neuropathology, immunohistochemical analysis of neuronal vulnerability, and comparison with related neurodegenerative disorders. *Acta neuropathologica,* 88**,** 397-404.

HOGAN, B., COSTANTINI, F. & LACY, E. 1986. *Manipulating the mouse embryo: a laboratory manual*, Cold spring harbor laboratory Cold Spring Harbor, NY.

HOLZBAUR, E. L. 2004. Motor neurons rely on motor proteins. *Trends in cell biology,* 14**,** 233-240.

HUANG, C., ZHOU, H., TONG, J., CHEN, H., LIU, Y.-J., WANG, D., WEI, X. & XIA, X.-G. 2011. FUS transgenic rats develop the phenotypes of amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *PLoS Genet,* 7**,** e1002011.
HUTTON, M., LENDON, C. L., RIZZU, P., BAKER, M., FROELICH, S., HOULDEN, H., PICKERING-BROWN, S., CHAKRAVERTY, S., ISAACS, A. & GROVER, A. 1998. Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature,* 393**,** 702-705.
IMMANUEL, D., ZINSZNER, H. & RON, D. 1995. Association of SARFH (sarcoma-associated RNA-binding fly homolog) with regions of chromatin transcribed by RNA polymerase II. *Molecular and cellular biology,* 15**,** 4562-4571.
ITO, D., SEKI, M., TSUNODA, Y., UCHIYAMA, H. & SUZUKI, N. 2011. Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. *Annals of neurology,* 69**,** 152-162.
JANKOVIC, J. 2008. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry,* 79**,** 368-376.
KABASHI, E., VALDMANIS, P. N., DION, P., SPIEGELMAN, D., MCCONKEY, B. J., VELDE, C. V., BOUCHARD, J.-P., LACOMBLEZ, L., POCHIGAEVA, K. & SALACHAS, F. 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature genetics,* 40**,** 572-574.
KAER, K. & SPEEK, M. 2013. Retroelements in human disease. *Gene,* 518**,** 231-241.
KANDEL, E. R., SCHWARTZ, J. H., JESSELL, T. M., SIEGELBAUM, S. A. & HUDSPETH, A. 2000. *Principles of neural science,* McGraw-hill New York.
KANG, H., YI, Y., KIM, H., HONG, Y., SEONG, Y. & BAE, I. 2012. BRCA1 negatively regulates IGF-1 expression through an estrogen-responsive element-like site. *Cell death & disease,* 3**,** e336.
KAVANAGH, D. H., DWYER, S., O'DONOVAN, M. C. & OWEN, M. J. 2013. The ENCODE project: implications for psychiatric genetics. *Mol Psychiatry*.
KAYE, A. D. & DAVIS, S. F. 2014. *Principles of Neurophysiological Assessment, Mapping, and Monitoring,* Springer.
KELLY, E. B. 2012. *Encyclopedia of Human Genetics and Disease,* ABC-CLIO.
KENNETH, S. S. 2003. Anatomy and physiology: The unity of form and function. USA: McGraw-Hill.
KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome research,* 12**,** 996-1006.
KHOKHA, M. K. & LOOTS, G. G. 2005. Strategies for characterising cis-regulatory elements in Xenopus. *Briefings in functional genomics & proteomics,* 4**,** 58-68.
KHURSHEED, K., WILM, T. P., CASHMAN, C., QUINN, J. P., BUBB, V. J. & MOSS, D. J. 2015. Characterisation of multiple regulatory domains spanning the major transcriptional start site of the FUS gene, a candidate gene for motor neurone disease. *Brain research,* 1595**,** 1-9.
KIM, V. N., HAN, J. & SIOMI, M. C. 2009. Biogenesis of small RNAs in animals. *Nature reviews Molecular cell biology,* 10**,** 126-139.
KONDOH, H., UCHIKAWA, M. & KAMACHI, Y. 2004. Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation. *International Journal of Developmental Biology,* 48**,** 819-827.
KONKEL, M. K. & BATZER, M. A. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Seminars in cancer biology, 2010. Elsevier, 211-221.
KRISTJUHAN, A. & SVEJSTRUP, J. Q. 2004. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. *The EMBO journal,* 23**,** 4243-4252.
KRUEGER, K. C. 2009. *Transcriptional regulation of FEV, a human serotonin neuron developmental control gene.* Case Western Reserve University.

KRYNDUSHKIN, D., WICKNER, R. B. & SHEWMAKER, F. 2011. FUS/TLS forms cytoplasmic aggregates, inhibits cell growth and interacts with TDP-43 in a yeast model of amyotrophic lateral sclerosis. *Protein & cell,* 2**,** 223-236.

KWIATKOWSKI, T. J., BOSCO, D., LECLERC, A., TAMRAZIAN, E., VANDERBURG, C., RUSS, C., DAVIS, A., GILCHRIST, J., KASARSKIS, E. & MUNSAT, T. 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science,* 323**,** 1205-1208.

LAGIER-TOURENNE, C. & CLEVELAND, D. W. 2009. Rethinking als: The fus about tdp-43. *Cell,* 136**,** 1001-1004.

LANDERS, J. E., MELKI, J., MEININGER, V., GLASS, J. D., VAN DEN BERG, L. H., VAN ES, M. A., SAPP, P. C., VAN VUGHT, P. W., MCKENNA-YASEK, D. M. & BLAUW, H. M. 2009. Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences,* 106**,** 9004-9009.

LAW, W. J., CANN, K. L. & HICKS, G. G. 2006. TLS, EWS and TAF15: a model for transcriptional integration of gene expression. *Briefings in functional genomics & proteomics,* 5**,** 8-14.

LE NAOUR, F., MISEK, D. E., KRAUSE, M. C., DENEUX, L., GIORDANO, T. J., SCHOLL, S. & HANASH, S. M. 2001. Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin Cancer Res,* 7**,** 3328-35.

LEIN, E. S., HAWRYLYCZ, M. J., AO, N., AYRES, M., BENSINGER, A., BERNARD, A., BOE, A. F., BOGUSKI, M. S., BROCKWAY, K. S. & BYRNES, E. J. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature,* 445**,** 168-176.

LIEM, K. F., TREMML, G., ROELINK, H. & JESSELL, T. M. 1995. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell,* 82**,** 969-979.

LOOTS, G. & OVCHARENKO, I. 2007. ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics,* 23**,** 122-124.

LUDOLPH, A., DRORY, V., HARDIMAN, O., NAKANO, I., RAVITS, J., ROBBERECHT, W. & SHEFNER, J. 2015. A revision of the El Escorial criteria-2015. *Amyotroph Lateral Scler Frontotemporal Degener,* 16**,** 291-292.

MACEDO, M. G., ANAR, B., BRONNER, I. F., CANNELLA, M., SQUITIERI, F., BONIFATI, V., HOOGEVEEN, A., HEUTINK, P. & RIZZU, P. 2003. The DJ-1L166P mutant protein associated with early onset Parkinson's disease is unstable and forms higher-order protein complexes. *Human molecular genetics,* 12**,** 2807-2816.

MACKEIGAN, J. P., CLEMENTS, C. M., LICH, J. D., POPE, R. M., HOD, Y. & TING, J. P. 2003a. Proteomic Profiling Drug-Induced Apoptosis in Non-Small Cell Lung Carcinoma Identification of RS/DJ-1 and RhoGDIα. *Cancer research,* 63**,** 6928-6934.

MACKEIGAN, J. P., CLEMENTS, C. M., LICH, J. D., POPE, R. M., HOD, Y. & TING, J. P. 2003b. Proteomic profiling drug-induced apoptosis in non-small cell lung carcinoma: identification of RS/DJ-1 and RhoGDIalpha. *Cancer Res,* 63**,** 6928-34.

MACKENZIE, A. & QUINN, J. 1999. A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. *Proceedings of the National Academy of Sciences,* 96**,** 15251-15255.

MACKENZIE, A. & QUINN, J. P. 2004. Post-genomic approaches to exploring neuropeptide gene mis-expression in disease. *Neuropeptides,* 38**,** 1-15.

MACKENZIE, I. R., RADEMAKERS, R. & NEUMANN, M. 2010. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *The Lancet Neurology,* 9**,** 995-1007.

MATSUMATA, M., UCHIKAWA, M., KAMACHI, Y. & KONDOH, H. 2005. Multiple N-cadherin enhancers identified by systematic functional screening indicate its Group B1 SOX-dependent regulation in neural and placodal development. *Developmental biology,* 286**,** 601-617.

MCCLINTOCK, B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences,* 36**,** 344-355.

MIKI, H., SETOU, M., KANESHIRO, K. & HIROKAWA, N. 2001. All kinesin superfamily protein, KIF, genes in mouse and human. *Proceedings of the National Academy of Sciences,* 98**,** 7004-7011.

MITCHELL, J. C., MCGOLDRICK, P., VANCE, C., HORTOBAGYI, T., SREEDHARAN, J., ROGELJ, B., TUDOR, E. L., SMITH, B. N., KLASEN, C. & MILLER, C. C. 2013. Overexpression of human wild-type FUS causes progressive motor neuron degeneration in an age-and dose-dependent fashion. *Acta neuropathologica,* 125**,** 273-288.

MIZIELINSKA, S., RIDLER, C. E., BALENDRA, R., THOENG, A., WOODLING, N. S., GRÄSSER, F. A., PLAGNOL, V., LASHLEY, T., PARTRIDGE, L. & ISAACS, A. M. 2017. Bidirectional nucleolar dysfunction in C9orf72 frontotemporal lobar degeneration. *Acta neuropathologica communications,* 5**,** 29.

MOROHOSHI, F., OOTSUKA, Y., ARAI, K., ICHIKAWA, H., MITANI, S., MUNAKATA, N. & OHKI, M. 1998. Genomic structure of the human RBP56/hTAFII68 and FUS/TLS genes. *Gene,* 221**,** 191-198.

MUNAFÒ, M. R. & JOHNSTONE, E. C. 2008. GENETIC STUDY: Smoking status moderates the association of the dopamine D4 receptor (DRD4) gene VNTR polymorphism with selective processing of smoking-related cues. *Addiction biology,* 13**,** 435-439.

MURAMATSU, T., MIZUTANI, Y., OHMORI, Y. & OKUMURA, J.-I. 1997. Comparison of three nonviral transfection methods for foreign gene expression in early chicken embryosin ovo. *Biochemical and biophysical research communications,* 230**,** 376-380.

MUTIHAC, R., ALEGRE-ABARRATEGUI, J., GORDON, D., FARRIMOND, L., YAMASAKI-MANN, M., TALBOT, K. & WADE-MARTINS, R. 2015. TARDBP pathogenic mutations increase cytoplasmic translocation of TDP-43 and cause reduction of endoplasmic reticulum Ca 2+ signaling in motor neurons. *Neurobiology of disease,* 75**,** 64-77.

NAGAKUBO, D., TAIRA, T., KITAURA, H., IKEDA, M., TAMAI, K., IGUCHI-ARIGA, S. M. & ARIGA, H. 1997. DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation withras. *Biochemical and biophysical research communications,* 231**,** 509-513.

NAKANO, T., WINDREM, M., ZAPPAVIGNA, V. & GOLDMAN, S. A. 2005. Identification of a conserved 125 base-pair Hb9 enhancer that specifies gene expression to spinal motor neurons. *Developmental biology,* 283**,** 474-485.

NEUMANN, E., GERISCH, G. & OPATZ, K. 1980. Cell fusion induced by high electric impulses applied to Dictyostelium. *Naturwissenschaften,* 67**,** 414-415.

NEUMANN, M., RADEMAKERS, R., ROEBER, S., BAKER, M., KRETZSCHMAR, H. A. & MACKENZIE, I. R. 2009. A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain,* 132**,** 2922-2931.

NEUMANN, M., SAMPATHU, D. M., KWONG, L. K., TRUAX, A. C., MICSENYI, M. C., CHOU, T. T., BRUCE, J., SCHUCK, T., GROSSMAN, M. & CLARK, C. M. 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science,* 314**,** 130-133.

ONG, K. K., PHILLIPS, D. I., FALL, C., POULTON, J., BENNETT, S. T., GOLDING, J., TODD, J. A. & DUNGER, D. B. 1999. The insulin gene VNTR, type 2 diabetes and birth weight. *Nature genetics,* 21**,** 262-263.

OROZCO, D., TAHIROVIC, S., RENTZSCH, K., SCHWENK, B. M., HAASS, C. & EDBAUER, D. 2012. Loss of fused in sarcoma (FUS) promotes pathological Tau splicing. *EMBO reports,* 13**,** 759-764.

OVCHARENKO, I., NOBREGA, M. A., LOOTS, G. G. & STUBBS, L. 2004. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic acids research,* 32**,** W280-W286.

PAREDES, U. M., BUBB, V. J., HADDLEY, K., MACHO, G. A. & QUINN, J. P. 2011. An evolutionary conserved region (ECR) in the human dopamine receptor D4 gene supports reporter gene expression in primary cultures derived from the rat cortex. *BMC neuroscience,* 12**,** 46.

PAVLICEK, A., GENTLES, A. J., PAČES, J., PAČES, V. & JURKA, J. 2006. Retroposition of processed pseudogenes: the impact of RNA stability and translational control. *Trends in Genetics,* 22**,** 69-73.

PENNACCHIO, L. A., AHITUV, N., MOSES, A. M., PRABHAKAR, S., NOBREGA, M. A., SHOUKRY, M., MINOVITSKY, S., DUBCHAK, I., HOLT, A. & LEWIS, K. D. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature,* 444**,** 499-502.

PENNACCHIO, L. A. & RUBIN, E. M. 2001. Genomic strategies to identify mammalian regulatory sequences. *Nature reviews genetics,* 2**,** 100-109.

PÉREZ-MANCERA, P. A. & SÁNCHEZ-GARÍA, I. 2004. FUS (fusion involved in t (12; 16) in malignant liposarcoma).

PIERANI, A., BRENNER-MORTON, S., CHIANG, C. & JESSELL, T. M. 1999. A sonic hedgehog–independent, retinoid-activated pathway of neurogenesis in the ventral spinal cord. *Cell,* 97**,** 903-915.

PLACZEK, M., TESSIER-LAVIGNE, M., YAMADA, T., JESSELL, T. & DODD, J. 1990. Mesodermal control of neural cell identity: floor plate induction by the notochord. *Science,* 250**,** 985.

POWERS, C. A., MATHUR, M., RAAKA, B. M., RON, D. & SAMUELS, H. H. 1998. TLS (translocated-in-liposarcoma) is a high-affinity interactor for steroid, thyroid hormone, and retinoid receptors. *Molecular endocrinology,* 12**,** 4-18.

PULS, I., JONNAKUTY, C., LAMONTE, B. H., HOLZBAUR, E. L., TOKITO, M., MANN, E., FLOETER, M. K., BIDUS, K., DRAYNA, D. & OH, S. J. 2003. Mutant dynactin in motor neuron disease. *Nature genetics,* 33**,** 455-456.

PULVERS, J. N. & HUTTNER, W. B. 2009. Brca1 is required for embryonic development of the mouse cerebral cortex to normal size by preventing apoptosis of early neural progenitors. *Development,* 136**,** 1859-1868.

PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., KATZ, L. C., LAMANTIA, A.-S., MCNAMARA, J. O. & WILLIAMS, S. M. 2001. Neuroscience. *Sunderland, MA: Sinauer Associates,* 3.

RABBITTS, T., FORSTER, A., LARSON, R. & NATHAN, P. 1993. Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t (12; 16) in malignant liposarcoma. *Nature genetics,* 4**,** 175-180.

RAO, M. V., MOHAN, P. S., KUMAR, A., YUAN, A., MONTAGNA, L., CAMPBELL, J., ESPREAFICO, E. M., JULIEN, J. P. & NIXON, R. A. 2011. The myosin Va head domain binds to the neurofilament-L rod and modulates endoplasmic reticulum (ER) content and distribution within axons. *PloS one,* 6**,** e17087-e17087.

REID, E., KLOOS, M., ASHLEY-KOCH, A., HUGHES, L., BEVAN, S., SVENSON, I. K., GRAHAM, F. L., GASKELL, P. C., DEARLOVE, A. & PERICAK-VANCE, M. A. 2002. A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10). *The American Journal of Human Genetics,* 71**,** 1189-1194.

RICHARDSON, S. R., MORELL, S. & FAULKNER, G. J. 2014. L1 retrotransposons and somatic mosaicism in the brain. *Annual review of genetics,* 48**,** 1-27.

RIKU, Y., ATSUTA, N., YOSHIDA, M., TATSUMI, S., IWASAKI, Y., MIMURO, M., WATANABE, H., ITO, M., SENDA, J. & NAKAMURA, R. 2014. Differential

motor neuron involvement in progressive muscular atrophy: a comparative study with amyotrophic lateral sclerosis. *BMJ open,* 4**,** e005213.
ROBERTS, J., SCOTT, A. C., HOWARD, M. R., BREEN, G., BUBB, V. J., KLENOVA, E. & QUINN, J. P. 2007. Differential regulation of the serotonin transporter gene by lithium is mediated by transcription factors, CCCTC binding protein and Y-box binding protein 1, through the polymorphic intron 2 variable number tandem repeat. *The Journal of Neuroscience,* 27**,** 2793-2801.
ROSEN, D. R., SIDDIQUEF, T., PATTERSON, D., FIGLEWICZ, D. A., SAPP, P., HENTATIF, A., O'REGAN, J. P., DENGF, H.-X., MCKENNA-YASEK, D. & CAYABYABI, A. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic. *Nature,* 362**,** 4.
RYTTSEN, F., FARRE, C., BRENNAN, C., WEBER, S. G., NOLKRANTZ, K., JARDEMARK, K., CHIU, D. T. & ORWAR, O. 2000. Characterization of single-cell electroporation by using patch-clamp and fluorescence microscopy. *Biophysical Journal,* 79**,** 1993-2001.
SABATELLI, M., CONTE, A. & ZOLLINO, M. 2013a. Clinical and genetic heterogeneity of amyotrophic lateral sclerosis. *Clinical Genetics,* 83**,** 408-416.
SABATELLI, M., MONCADA, A., CONTE, A., LATTANTE, S., MARANGI, G., LUIGETTI, M., LUCCHINI, M., MIRABELLA, M., ROMANO, A. & DEL GRANDE, A. 2013b. Mutations in the 3′ untranslated region of FUS causing FUS overexpression are associated with amyotrophic lateral sclerosis. *Human molecular genetics,* 22**,** 4748-4755.
SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular cloning: a laboratory manual*, Cold spring harbor laboratory press.
SANDER, M., PAYDAR, S., ERICSON, J., BRISCOE, J., BERBER, E., GERMAN, M., JESSELL, T. M. & RUBENSTEIN, J. L. 2000. Ventral neural patterning by Nkx homeobox genes: Nkx6. 1 controls somatic motor neuron and ventral interneuron fates. *Genes & development,* 14**,** 2134-2139.
SAVAGE, A., WILM, T., KHURSHEED, K., SHATUNOV, A. & MORRISON, K. 2014a. An Evaluation of a SVA Retrotransposon in the FUS Promoter as a.
SAVAGE, A. L., BUBB, V. J., BREEN, G. & QUINN, J. P. 2013a. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC evolutionary biology,* 13**,** 1.
SAVAGE, A. L., BUBB, V. J., BREEN, G. & QUINN, J. P. 2013b. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC evolutionary biology,* 13**,** 101.
SAVAGE, A. L., WILM, T. P., KHURSHEED, K., SHATUNOV, A., MORRISON, K. E., SHAW, P. J., SHAW, C. E., SMITH, B., BREEN, G. & AL-CHALABI, A. 2014b. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS One,* 9**,** e90833.
SHERIDAN, J. D. 1968. Electrophysiological evidence for low-resistance intercellular junctions in the early chick embryo. *The Journal of cell biology,* 37**,** 650-659.
SIMPSON, C. L., LEMMENS, R., MISKIEWICZ, K., BROOM, W. J., HANSEN, V. K., VAN VUGHT, P. W., LANDERS, J. E., SAPP, P., VAN DEN BOSCH, L. & KNIGHT, J. 2008. Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. *Human molecular genetics,* 18**,** 472-481.
SIMPSON, C. L., LEMMENS, R., MISKIEWICZ, K., BROOM, W. J., HANSEN, V. K., VAN VUGHT, P. W., LANDERS, J. E., SAPP, P., VAN DEN BOSCH, L. & KNIGHT, J. 2009. Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. *Human molecular genetics,* 18**,** 472-481.
SINGH, Y., CHEN, H., ZHOU, Y., FÖLLER, M., MAK, T. W., SALKER, M. S. & LANG, F. 2015. Differential effect of DJ-1/PARK7 on development of natural and induced regulatory T cells. *Scientific reports,* 5**,** 17723.

SPENCER, P. S. & PALMER, V. S. 2012. Interrelationships of undernutrition and neurotoxicity: food for thought and research attention. *Neurotoxicology,* 33**,** 605-616.

SPROVIERO, W., LA BELLA, V., MAZZEI, R., VALENTINO, P., RODOLICO, C., SIMONE, I. L., LOGROSCINO, G., UNGARO, C., MAGARIELLO, A. & PATITUCCI, A. 2012. FUS mutations in sporadic amyotrophic lateral sclerosis: clinical and genetic analysis. *Neurobiology of aging,* 33**,** 837. e1-837. e5.

SREEDHARAN, J., BLAIR, I. P., TRIPATHI, V. B., HU, X., VANCE, C., ROGELJ, B., ACKERLEY, S., DURNALL, J. C., WILLIAMS, K. L. & BURATTI, E. 2008. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science,* 319**,** 1668-1672.

STEELE, J. C., RICHARDSON, J. C. & OLSZEWSKI, J. Progressive supranuclear palsy: a heterogeneous degeneration involving the brain stem, Basal Ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Seminars in neurology, 2014. 129.

STIFANI, N. 2014. Motor neurons and the generation of spinal motor neuron diversity. *Frontiers in cellular neuroscience,* 8.

STRACK, R. L., STRONGIN, D. E., BHATTACHARYYA, D., TAO, W., BERMAN, A., BROXMEYER, H. E., KEENAN, R. J. & GLICK, B. S. A noncytotoxic DsRed variant for whole-cell labeling. SPIE BiOS: Biomedical Optics, 2009. International Society for Optics and Photonics, 719109-719109-12.

STROMINGER, N. L., DEMAREST, R. J. & LAEMLE, L. B. 2012. *Noback's human nervous system: structure and function,* Springer Science & Business Media.

SZPAKOWSKI, S., SUN, X., LAGE, J. M., DYER, A., RUBINSTEIN, J., KOWALSKI, D., SASAKI, C., COSTA, J. & LIZARDI, P. M. 2009. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. *Gene,* 448**,** 151-167.

TAKAHASHI, K., TAIRA, T., NIKI, T., SEINO, C., IGUCHI-ARIGA, S. M. & ARIGA, H. 2001. DJ-1 positively regulates the androgen receptor by impairing the binding of PIASxα to the receptor. *Journal of Biological Chemistry,* 276**,** 37556-37563.

TALBOT, K. 2002. Motor neurone disease. *Postgraduate medical journal,* 78**,** 513-519.

TALBOT, K. & MARSDEN, R. 2008. *Motor neuron disease,* Oxford University Press.

TANABE, Y., WILLIAM, C. & JESSELL, T. M. 1998. Specification of motor neuron identity by the MNR2 homeodomain protein. *Cell,* 95**,** 67-80.

TANAKA, J., HARADA, H., ITO, K., OGURA, T. & NAKAMURA, H. 2010. Gene manipulation of chick embryos in vitro, early chick culture, and long survival in transplanted eggs. *Development, growth & differentiation,* 52**,** 629-634.

TANNER, N. K. & LINDER, P. 2001. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Molecular cell,* 8**,** 251-262.

TEISSIE, J. & ROLS, M.-P. 1993. An experimental evaluation of the critical potential difference inducing cell membrane electropermeabilization. *Biophysical journal,* 65**,** 409.

THAKUR, G. S., DAIGLE JR, B. J., DEAN, K. R., ZHANG, Y., RODRIGUEZ-FERNANDEZ, M., HAMMAMIEH, R., YANG, R., JETT, M., PALMA, J. & PETZOLD, L. R. 2015. Systems biology approach to understanding post-traumatic stress disorder. *Molecular BioSystems,* 11**,** 980-993.

TIMMER, J., JOHNSON, J. & NISWANDER, L. 2001. The use of in ovo electroporation for the rapid analysis of neural-specific murine enhancers. *Genesis,* 29**,** 123-132.

TROJSI, F., MONSURRÒ, M. R. & TEDESCHI, G. 2013. Exposure to environmental toxicants and pathogenesis of amyotrophic lateral sclerosis: state of the art and research perspectives. *International journal of molecular sciences,* 14**,** 15286-15311.

TURNER, M. R., HARDIMAN, O., BENATAR, M., BROOKS, B. R., CHIO, A., DE CARVALHO, M., INCE, P. G., LIN, C., MILLER, R. G. & MITSUMOTO, H. 2013. Controversies and priorities in amyotrophic lateral sclerosis. *The Lancet Neurology,* 12**,** 310-322.

UCHIKAWA, M. 2008. Enhancer analysis by chicken embryo electroporation with aid of genome comparison. *Development, growth & differentiation,* 50**,** 467-474.
UCHIKAWA, M., ISHIDA, Y., TAKEMOTO, T., KAMACHI, Y. & KONDOH, H. 2003. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Developmental cell,* 4**,** 509-519.
UCHIKAWA, M., TAKEMOTO, T., KAMACHI, Y. & KONDOH, H. 2004. Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison. *Mechanisms of development,* 121**,** 1145-1158.
ULLU, E., MURPHY, S. & MELLI, M. 1982. Human 7SL RNA consists of a 140 nucleotide middle-repetitive sequence inserted in an alu sequence. *Cell,* 29**,** 195-202.
URANISHI, H., TETSUKA, T., YAMASHITA, M., ASAMITSU, K., SHIMIZU, M., ITOH, M. & OKAMOTO, T. 2001. Involvement of the pro-oncoprotein TLS (translocated in liposarcoma) in nuclear factor-κB p65-mediated transcription as a coactivator. *Journal of Biological Chemistry,* 276**,** 13395-13401.
URWIN, H., JOSEPHS, K. A., ROHRER, J. D., MACKENZIE, I. R., NEUMANN, M., AUTHIER, A., SEELAAR, H., VAN SWIETEN, J. C., BROWN, J. M. & JOHANNSEN, P. 2010. FUS pathology defines the majority of tau-and TDP-43-negative frontotemporal lobar degeneration. *Acta neuropathologica,* 120**,** 33-41.
VAN DEERLIN, V. M., LEVERENZ, J. B., BEKRIS, L. M., BIRD, T. D., YUAN, W., ELMAN, L. B., CLAY, D., WOOD, E. M., CHEN-PLOTKIN, A. S. & MARTINEZ-LAGE, M. 2008. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *The Lancet Neurology,* 7**,** 409-416.
VAN MOSSEVELDE, S., VAN DER ZEE, J., GIJSELINCK, I., SLEEGERS, K., DE BLEECKER, J., SIEBEN, A., VANDENBERGHE, R., VAN LANGENHOVE, T., BAETS, J. & DERYCK, O. 2017. Clinical Evidence of Disease Anticipation in Families Segregating a C9orf72 Repeat Expansion. *Jama neurology,* 74**,** 445-452.
VANCE, C., SCOTTER, E. L., NISHIMURA, A. L., TROAKES, C., MITCHELL, J. C., KATHE, C., URWIN, H., MANSER, C., MILLER, C. C. & HORTOBÁGYI, T. 2013. ALS mutant FUS disrupts nuclear localisation and sequesters wild-type FUS within cytoplasmic stress granules. *Human molecular genetics***,** ddt117.
VASIEVA, O., CETINER, S., SAVAGE, A., SCHUMANN, G. G., BUBB, V. J. & QUINN, J. P. 2016. Primate specific retrotransposons, SVAs, in the evolution of networks that alter brain function. *arXiv preprint arXiv:1602.07642*.
VASILIOU, S. A., ALI, F. R., HADDLEY, K., CARDOSO, M. C., BUBB, V. J. & QUINN, J. P. 2012. The SLC6A4 VNTR genotype determines transcription factor binding and epigenetic variation of this gene in response to cocaine in vitro. *Addiction biology,* 17**,** 156-170.
VEJJAJIVA, A., FOSTER, J. & MILLER, H. 1967. Motor neuron disease: a clinical study. *Journal of the Neurological Sciences,* 4**,** 299-314.
WAIBEL, S., NEUMANN, M., RABE, M., MEYER, T. & LUDOLPH, A. 2010. Novel missense and truncating mutations in FUS/TLS in familial ALS. *Neurology,* 75**,** 815-817.
WANG, H., XING, J., GROVER, D., HEDGES, D. J., HAN, K., WALKER, J. A. & BATZER, M. A. 2005. SVA elements: a hominid-specific retroposon family. *Journal of molecular biology,* 354**,** 994-1007.
WEAVER, J. C. 1995. Electroporation theory. *Animal Cell Electroporation and Electrofusion Protocols***,** 3-28.
WEINGARTEN, M. D., LOCKWOOD, A. H., HWO, S.-Y. & KIRSCHNER, M. W. 1975. A protein factor essential for microtubule assembly. *Proceedings of the National Academy of Sciences,* 72**,** 1858-1862.
XIA, C.-H., ROBERTS, E. A., HER, L.-S., LIU, X., WILLIAMS, D. S., CLEVELAND, D. W. & GOLDSTEIN, L. S. 2003. Abnormal neurofilament transport caused by

targeted disruption of neuronal kinesin heavy chain KIF5A. *The Journal of cell biology,* 161**,** 55-66.

XING, J., ZHANG, Y., HAN, K., SALEM, A. H., SEN, S. K., HUFF, C. D., ZHOU, Q., KIRKNESS, E. F., LEVY, S. & BATZER, M. A. 2009. Mobile elements create structural variation: analysis of a complete human genome. *Genome research,* 19**,** 1516-1526.

YAMADA, M., IWATSUBO, T., MIZUNO, Y. & MOCHIZUKI, H. 2004. Overexpression of α-synuclein in rat substantia nigra results in loss of dopaminergic neurons, phosphorylation of α-synuclein and activation of caspase-9: resemblance to pathogenetic changes in Parkinson's disease. *Journal of neurochemistry,* 91**,** 451-461.

YANEZA, M., GILTHORPE, J. D., LUMSDEN, A. & TUCKER, A. S. 2002. No evidence for ventrally migrating neural tube cells from the mid-and hindbrain. *Developmental dynamics,* 223**,** 163-167.

YASUDA, K., MOMOSE, T. & TAKAHASHI, Y. 2000. Applications of microelectroporation for studies of chick embryogenesis. *Development, growth & differentiation,* 42**,** 203-206.

YUAN, A., RAO, M. V., SASAKI, T., CHEN, Y., KUMAR, A., LIEM, R. K., EYER, J., PETERSON, A. C., JULIEN, J.-P. & NIXON, R. A. 2006. α-Internexin is structurally and functionally associated with the neurofilament triplet proteins in the mature CNS. *The Journal of neuroscience,* 26**,** 10006-10019.

ZINSZNER, H., ALBALAT, R. & RON, D. 1994. A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. *Genes & development,* 8**,** 2513-2526.